

CAMILA RIBEIRO DA SILVA

**GRÁFICO DE CONTROLE DO TIPO SHEWHART CONSIDERANDO A
DISTRIBUIÇÃO GAMA UNITÁRIA INFLACIONADA**

RECIFE - PE

2023



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**Gráfico de Controle do tipo Shewhart considerando a distribuição gama
unitária inflacionada**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como requisito parcial para a obtenção do título de Doutora em Biometria e Estatística Aplicada.

Área de concentração: Biometria e Estatística Aplicada.

Orientador: Prof. Dr. Paulo José Duarte Neto

Coorientadores: Prof. Dr. Luiz Medeiros de A. Lima Filho e Prof. Dra. Tarciana Liberal Pereira

RECIFE - PE

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- S586g Silva, Camila Ribeiro da
Gráfico de Controle do tipo Shewhart considerando a distribuição gama unitária inflacionada / Camila Ribeiro da Silva. - 2023.
70 f. : il.
- Orientador: Paulo Jose Duarte Neto.
Coorientador: Luiz Medeiros de Araujo Lima Filho.
Inclui referências e apêndice(s).
- Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2023.
1. Gráficos de controle. 2. Dados inflacionados. 3. Distribuição gama unitária inflacionada. 4. Taxas e proporções. I. Neto, Paulo Jose Duarte, orient. II. Filho, Luiz Medeiros de Araujo Lima, coorient. III. Título

CAMILA RIBEIRO DA SILVA

Gráfico de Controle do tipo Shewhart considerando a distribuição gama unitária inflacionada

Tese para obtenção do título de Doutora em Biometria e Estatística Aplicada, defendida e julgada em 25/08/2023.

Orientadores:

Prof. Dr. Paulo José Duarte Neto - Orientador
Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Luiz Medeiros de Araujo Lima Filho - Coorientador
Universidade Federal da Paraíba - UFPB

Profa. Dra. Tarciana Liberal Pereira - Coorientadora
Universidade Federal da Paraíba - UFPB

Banca Examinadora:

Prof. Dr. Fábio Mariano Bayer
Universidade Federal de Santa Maria - UFSM

Prof. Dr. Josimar Mendes de Vasconcelos
Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Lucian Bogdan Bejan
Universidade Federal Rural de Pernambuco - UFRPE

Profa. Dra. Tatiene Correia de Souza
Universidade Federal da Paraíba - UFPB

Dedico este trabalho aos meus pais, Rizete Maria e José, e ao meu esposo Alisson, com todo meu amor.

AGRADECIMENTOS

Agradeço a Deus por todas as bênçãos concedidas na minha vida.

À minha família, em especial, aos meus pais, por serem inspiração de força e dedicação. Por não medirem esforços para me apoiar e incentivar. Obrigada por tudo.

Ao meu esposo, Alisson, por quem tenho amor, respeito e admiração. Obrigada pelo incentivo e apoio em todos os momentos. Ao meu pequeno Miguel, que ainda não imagina, mas é minha fonte de motivação diária.

Aos meus orientadores, professor Paulo Duarte, Luiz Medeiros e Tarciana Liberal, por toda orientação, competência e amizade ao longo desse tempo. Obrigada por todo conhecimento transmitido e direcionamentos para a conclusão desta tese.

Aos amigos que fiz ao longo do curso, obrigada pelos momentos de descontração e partilha de conhecimento. Em especial, a Elielma, Catiane, Natália, Marciele, Lucas, João Rocha e Edvaldo, pela amizade e troca de experiências. Foi muito gratificante conviver com vocês.

Agradeço a banca examinadora, antecipadamente, pelas contribuições dadas para a melhoria deste trabalho desde a qualificação. Especialmente, as professoras Tatiene Correia e Patrícia Ospina, vocês fazem parte da minha trajetória acadêmica de uma forma especial.

A todos os Professores do Programa de Pós-graduação em Biometria e Estatística Aplicada, por contribuírem para minha formação acadêmica.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo apoio financeiro.

A persistência é o menor caminho do êxito.
(Charles Chaplin)

RESUMO

O controle estatístico de processos (CEP) é um dos eixos que contemplam a área de controle de qualidade. Os gráficos de controle, uma das ferramentas do CEP, foram originalmente desenvolvidos para monitorar processos industriais. Entretanto, nos últimos anos, sua aplicação tem se mostrado de grande relevância no monitoramento de variáveis em diferentes contextos. Em alguns casos, o objetivo é monitorar o comportamento de variáveis que assumem valores nos intervalos $[0,1)$ ou $(0,1]$, isto é, variáveis com inflação de zeros ou uns. Tendo em vista a limitação de gráficos de controle apropriados para monitorar variáveis nos intervalos $(0,1]$ e $[0,1)$, o respectivo trabalho de tese visa propor gráficos de controle baseados na distribuição gama unitária inflacionada. Dessa forma, inicialmente, propomos a distribuição gama unitária inflacionada em zero e um e derivamos suas principais propriedades. Adicionalmente, apresentamos expressões para a obtenção dos estimadores de máxima verossimilhança dos parâmetros e conduzimos simulações para avaliar a performance dos intervalos de confiança e testes de hipóteses, em amostras de tamanho finito. Na formulação do modelo, partimos de uma parametrização em que a distribuição proposta é expressa em termos da média da distribuição inflacionada. Essa parametrização torna-se mais atrativa, tendo em vista que em gráficos de controle quando o tamanho do subgrupo é maior que 1, é comum o interesse em monitorar a média do processo. Além disso, duas aplicações a dados reais são apresentadas para ilustrar a aplicabilidade da distribuição proposta. Na sequência, propomos o gráfico de controle gama unitária inflacionado para o monitoramento de variáveis que assumem os valores zero ou um. Na formulação do gráfico proposto, assumimos que a variável monitorada segue distribuição gama unitária inflacionada. Um extenso estudo de simulação de Monte Carlo foi realizado para avaliar a performance dos gráficos de controle em termos de comprimento da sequência. Realizamos uma comparação entre o gráfico de controle gama unitária inflacionado e o gráfico de controle beta inflacionado, a partir de duas abordagens. Na primeira, consideramos observações individuais e na segunda, subgrupos amostrais de tamanho $m = 8, 15, 30$ e 50 . Os resultados numéricos evidenciam que o gráfico proposto apresentou bom desempenho nas duas abordagens consideradas. Com base em aplicações a conjunto de dados reais, é possível verificar que o gráfico de controle proposto pode ser considerado uma boa alternativa para o monitoramento de variáveis definidas nos intervalos $(0,1]$ e $[0,1)$.

Palavras-chave: Gráficos de controle, dados inflacionados, distribuição gama unitária inflacionada, taxas e proporções.

ABSTRACT

Statistical process control (SPC) is one of the pillars of quality control. Control charts, one of the tools in SPC, were originally developed to monitor industrial processes. However, in recent years, its application has highly relevant in monitoring variables in different contexts. In some cases, the aim is to monitor the behavior of variables that take on values in the intervals $[0,1)$ or $(0,1]$, i.e. variables with an inflation of zeros or ones. In view of the limitation of control charts suitable for monitoring variables in the intervals $(0,1]$ and $[0,1)$, this thesis aims to propose control charts based on the inflated unit gamma distribution. We propose the unitary gamma distribution inflated at zero and one and derive its main properties. In addition, we present expressions for obtaining maximum likelihood estimators for the parameters and conduct simulations to evaluate the performance of confidence intervals and hypothesis tests on finite sample sizes. In formulating the model, we started with a parameterization in which the proposed distribution is expressed in terms of the mean of the inflated distribution. This parameterization is more attractive because in control charts, when the size of the subgroup is greater than 1, it is common to monitor the average of the process. In addition, two applications to real data are presented to illustrate the applicability of the proposed distribution. Next, we propose the inflated unit gamma control chart, for monitoring variables that take on the values zero or one. In formulating the proposed graph, we assume that the monitored variable follows an inflated unit gamma distribution. An extensive Monte Carlo simulation study was carried out to evaluate the performance of the control charts in terms of sequence length. We compared the inflated unit gamma control chart and the inflated beta control chart using two approaches. In the first, we considered individual observations and in the second, sample subgroups of size $m = 8, 15, 30$ and 50 . The numerical results show that the proposed graph performed well in the two approaches considered. Based on applications to real data sets, it can be verified that the proposed control chart can be considered a good alternative for monitoring variables defined in the intervals $(0,1]$ and $[0,1)$.

Keywords: Control charts, inflated data, inflated unit gamma distribution, rates and proportions .

SUMÁRIO

1	INTRODUÇÃO GERAL	12
2	DISTRIBUIÇÃO GAMA UNITÁRIA INFLACIONADA E SUA APLICAÇÃO NA MODELAGEM DE DADOS DE SANEAMENTO BÁSICO	14
2.1	Introdução	15
2.2	Distribuição gama unitária	17
2.2.1	Distribuição proposta	18
2.2.2	Estimação	21
2.2.3	Testes de hipóteses	22
2.3	Estudo de simulação	23
2.3.1	Resultados numéricos	24
2.4	Aplicações	34
2.4.1	Dados de acesso à rede de esgotamento sanitário para o Estado de São Paulo	34
2.4.2	Proporção de esgoto tratado em relação à água consumida no Estado de São Paulo	36
2.5	Conclusão	38
3	GRÁFICO DE CONTROLE DO TIPO SHEWHART CONSIDERANDO A DISTRIBUIÇÃO GAMA UNITÁRIA INFLACIONADA EM ZERO OU UM	40
3.1	Introdução	41
3.2	Gráfico de controle gama unitária inflacionado	42
3.2.1	Distribuição gama unitária inflacionada	43
3.2.2	Gráfico de controle para monitorar medidas individuais	44
3.2.3	Gráfico de controle \bar{X}	44
3.3	Avaliação de desempenho dos gráficos de controle	46
3.4	Medidas individuais	47
3.5	Observações não individuais	50
3.6	Aplicações	58

3.6.1	Dados de letalidade por dengue	58
3.6.2	Dados de umidade relativa	60
3.7	Conclusão	61
4	CONCLUSÃO GERAL	63
	REFERÊNCIAS	64
	APÊNDICE A –	70

1 INTRODUÇÃO GERAL

A área de modelagem estatística tem buscado o desenvolvimento de modelos cada vez mais realistas e que possibilitem a descrição de variáveis de diversos tipos. Os modelos estatísticos para explicar variáveis no intervalo unitário, como taxas ou proporções, foram propostos por alguns autores (FERRARI; CRIBARI-NETO, 2004; ESPINHEIRA; SILVA, 2020; PUMI et al., 2020). No entanto, em alguns casos, nos deparamos com variáveis que assumem, inclusive, os valores 0 e 1 com probabilidade positiva. Para que seja possível modelar tais variáveis, foram propostas distribuições mistas, denominadas de distribuições inflacionadas. Essas distribuições são resultado da composição entre duas distribuições, isto é, uma mistura entre uma distribuição contínua no intervalo $(0,1)$ e a distribuição de Bernoulli, cuja massa de probabilidade é concentrada em 0 ou 1. Vários autores apresentam a generalização de modelos contínuos para situações em que a variável aleatória assume valores nos intervalos $(0,1]$, $[0,1)$ ou $[0,1]$ (OSPINA; FERRARI, 2012; CRIBARI-NETO; SANTOS, 2019; CORPES, 2021; BAPAT; BHARDWAJ, 2021). Através dessas generalizações, é possível construir modelos que possam captar a massa de probabilidade concentrada nos pontos zero, um ou ambos, conforme o caso.

O controle de qualidade é uma área da estatística destinada a aplicação de métodos para o monitoramento de processos. Em algumas situações, a variável ou característica de qualidade monitorada pode estar distribuída nos intervalos $(0,1]$ ou $[0,1)$. Neste sentido, através dessas distribuições inflacionadas também é possível construir gráficos de controle para o monitoramento desses processos. Lima-Filho et al. (2019) propuseram o gráfico de controle beta inflacionado para monitorar conjuntos de dados limitados no intervalo unitário padrão. O gráfico de controle proposto é baseado no pressuposto de que os dados seguem distribuição beta inflacionada. Além disso, recentemente, Lima-Filho et al. (2023) propuseram um gráfico de controle para o monitoramento de variáveis ambientais inflacionadas em zero ou um, a partir da distribuição Kumaraswamy. Neste sentido, o desenvolvimento de gráficos de controle baseados em novas distribuições inflacionadas no intervalo unitário é de suma importância, uma vez que podem conferir maior flexibilidade no monitoramento de variáveis, principalmente em situações onde os gráficos já propostos podem não ser apropriados.

A distribuição gama unitária (GRASSIA, 1977) é uma distribuição contínua no intervalo unitário padrão $(0,1)$. Essa distribuição mostra-se bastante flexível, podendo assumir diversas formas (DEY et al., 2019). Alguns autores já exploraram essa distribuição e estudaram a mesma sob diversos aspectos. Mousa et al. (2016) apresentam o modelo de regressão para variáveis contínuas que tomam valores em um intervalo limitado, a partir da distribuição gama unitária. Mazucheli et al. (2018b) derivaram correções de viés de segunda ordem para os estimadores de máxima verossimilhança dos parâmetros. Diferentes métodos de estimação para os parâmetros da distribuição gama unitária são abordados por Dey et al. (2019). Ho et al. (2019) propuseram gráficos de controle tomando como base a distribuição gama unitária.

Tendo em vista a limitação de gráficos de controle apropriados para monitorar variáveis que assumem valores nos intervalos $(0,1]$ ou $[0,1)$ e a importância em monitorar essas variáveis, o respectivo trabalho de tese visa propor gráficos de controle baseados na distribuição gama unitária inflacionada. Neste sentido, inicialmente, propomos a distribuição gama unitária inflacionada em zero e/ou um e derivamos suas principais propriedades. Na formulação do modelo, partimos de uma parametrização similar a apresentada por Bayes e Valdivieso (2016), em que é possível modelar a média da distribuição inflacionada. Em seguida, propomos o gráfico de controle gama unitária inflacionado para o monitoramento de variáveis que assumem valores nos intervalos $(0,1]$ ou $[0,1)$. Na construção do gráfico proposto, assumimos que a variável monitorada segue distribuição gama unitária inflacionada. Para avaliar o desempenho do gráfico de controle proposto, utilizamos as medidas Average Runing Length - ARL, Median Run Length - MRL e Standard Deviation of Run Length - SDRL (MONTGOMERY, 2009; AHMAD et al., 2019; TEOH et al., 2017), bastante difundidas na literatura. Além disso, apresentamos duas aplicações a conjunto de dados reais para ilustrar a aplicação do gráfico proposto.

Os resultados apresentados a seguir (em formato de artigo) estão estruturados como segue. No segundo Capítulo apresentamos os resultados referentes ao primeiro artigo da tese. Ele é composto pela apresentação da distribuição gama unitária inflacionada, suas principais propriedades, estudos de simulação para avaliar estimadores pontuais e intervalares e a performance de testes de hipóteses. Adicionalmente, apresentamos duas aplicações a dados reais. Na sequência, terceiro Capítulo, apresentamos os resultados do segundo artigo, isto é, o gráfico de controle baseado na distribuição gama unitária inflacionada. Realizamos estudos de simulação para avaliar o desempenho do gráfico de controle proposto e comparamos os resultados com o gráfico de controle beta inflacionado (LIMA-FILHO et al., 2019). Por último, para ilustrar a aplicabilidade do gráfico proposto, apresentamos duas aplicações a banco de dados reais. No Capítulo 4, trazemos uma conclusão geral com o objetivo de sumarizar as principais contribuições do estudo e pontos importantes em cada uma das etapas.

2 DISTRIBUIÇÃO GAMA UNITÁRIA INFLACIONADA E SUA APLICAÇÃO NA MODELAGEM DE DADOS DE SANEAMENTO BÁSICO

RESUMO: A distribuição gama unitária é adequada para modelar variáveis no intervalo $(0,1)$. Em alguns cenários, no entanto, o conjunto de dados pode conter os valores zero e/ou um com probabilidade positiva. Neste caso, a modelagem é direcionada para variáveis restritas aos intervalos $(0,1]$, $[0,1)$ ou $[0,1]$, sendo necessário, portanto, considerar distribuições apropriadas. Neste artigo, propomos a distribuição gama unitária inflacionada e apresentamos suas principais propriedades. Derivamos expressões para a obtenção dos estimadores de máxima verossimilhança dos parâmetros da distribuição. Além disso, são conduzidas simulações para avaliar a performance dos intervalos de confiança e testes de hipóteses, em amostras de tamanho finito. Adicionalmente, apresentamos aplicações a dados reais para ilustrar o uso da distribuição proposta e realizamos uma comparação com outras distribuições inflacionadas da literatura. A distribuição gama unitária inflacionada apresenta resultados satisfatórios quanto ao ajuste de variáveis no contexto de saneamento básico e mostra-se uma potencial concorrente para as distribuições beta e Kumaraswamy inflacionadas. Além disso, algumas distribuições inflacionadas propostas na literatura estimam a média ou mediana considerando apenas a parte contínua do modelo, o que pode trazer algumas limitações práticas, ao contrário da distribuição proposta em que é possível estimar a média da distribuição inflacionada.

Palavras-chave: Dados inflacionados, distribuição gama unitária inflacionada, estimador de máxima verossimilhança, saneamento básico.

ABSTRACT: The unit gamma distribution is suitable for model variables in the interval $(0,1)$. In some scenarios, however, the dataset may contain values zero and/or one with probability positive. In this case, the modeling is directed to variables restricted to intervals $(0, 1]$, $[0, 1)$ or $[0,1]$, being necessary, therefore, to consider appropriate models. This paper proposes inflated unit gamma distribution and present its main properties. We derive expressions to obtain the maximum likelihood estimators of the distribution parameters. Additionally, simulations are conducted to evaluate the performance of confidence intervals and hypothesis tests, in finite sample sizes. Furthermore, we present applications to real data to illustrate the use of the proposed distribution and perform a comparison with other inflated distributions in the literature. The inflated unit gamma distribution presents satisfactory results regarding the adjustment of variables in the context of basic sanitation and proves to be a potential competitor for the inflated beta and Kumaraswamy distributions. Furthermore, some inflated distributions proposed in the literature estimate the mean or median considering only the continuous part of the model, which may bring some practical limitations, unlike the proposed distribution in which it is possible to estimate the mean of the inflated distribution.

Keywords: Inflated data, inflated unit gamma distribution, maximum likelihood estimator, basic sanitation.

2.1 INTRODUÇÃO

As distribuições de probabilidade podem ser caracterizadas como instrumentos de grande importância na descrição de fenômenos aleatórios. Atualmente, no campo da estatística, existe o interesse no desenvolvimento de novas classes de distribuições a fim de fornecer maior flexibilidade na modelagem de dados e, assim, englobar o maior número de fenômenos aleatórios. Algumas dessas novas distribuições, frequentemente, têm como característica principal serem resultado da composição entre duas ou mais distribuições e são denominadas de distribuições de mistura (YOO, 2004; OSPINA; FERRARI, 2010; RAHMAN et al., 2022; FERREIRA; MAZUCHELI, 2022).

A distribuição gama unitária (GU), proposta por Grassia (1977), ganhou relevante destaque na literatura ao longo dos anos. Ratnaparkhl e Mosimann (1990) usaram esta distribuição para derivar algumas novas distribuições tomando a GU como uma distribuição condicional. A distribuição GU também pode ser usada como uma alternativa para as distribuições beta (JOHNSON et al., 1995), simplex (BARNDORFF-NIELSEN; JØRGENSEN, 1991), Weibull unitária (MAZUCHELI et al., 2018a), Kumaraswamy (KUMARASWAMY, 1976) e Kumaraswamy modificada (SAGRILLO et al., 2021). Ratnaparkhl e Mosimann (1990) estudaram a transformação logarítmica e tipo lambda de Tukey na distribuição gama unitária. Por outro lado, Mousa et al. (2016) propuseram o modelo de regressão GU, ao passo que Mazucheli et al. (2018b) derivaram correções de viés de segunda ordem para os parâmetros dessa distribuição. Dey et al. (2019) realizaram comparações de métodos de estimação para a distribuição GU, enquanto Ho et al. (2019) consideraram a distribuição GU como base para a construção de gráficos de controle para monitorar taxas e proporções. Rocha et al. (2021) propuseram um resíduo baseado no vetor escore, além de medidas de influência local sob diferentes esquemas de perturbação para regressão gama unitária e modelaram a variável percentual de pessoas fora da linha de pobreza no cenário mundial. Os autores realizaram uma comparação com a distribuição beta e verificaram um melhor ajuste da distribuição GU. A distribuição GU também se mostra muito útil no campo aplicado, sendo utilizada para estimar a densidade de bactérias ou vírus em ensaios de diluição (GRASSIA, 1977).

A distribuição GU modela variáveis no intervalo $(0,1)$. No entanto, em alguns cenários, o conjunto de dados pode conter os valores zero e/ou um com probabilidade positiva e, assim, a distribuição GU não pode ser utilizada. Nesta situação, a modelagem é direcionada para variáveis que assumem valores nos intervalos $(0,1]$, $[0,1)$ ou $[0,1]$, sendo necessário analisar tais conjuntos de dados usando distribuições inflacionadas apropriadas. Neste contexto, Ospina e Ferrari (2010) introduziram a classe de distribuições beta inflacionadas, em que é permitido a presença de valores zero e/ou um nos dados. A palavra inflacionada sugere que a massa de probabilidade em alguns pontos excede o que é permitido pelo modelo proposto (TU, 2006). A ideia por trás de distribuições inflacionadas, no contexto de distribuições contínuas no intervalo $(0,1)$, é considerar que a distribuição dos dados é uma mistura entre uma distribuição contínua definida no intervalo $(0,1)$ e a distribuição de Bernoulli, a qual atribui probabilidades aos inteiros 0 e 1.

Algumas distribuições inflacionadas foram propostas na literatura com a finalidade de modelar variáveis dessa natureza. Por exemplo, Cribari-Neto e Santos (2019) apresentaram a distribuição Kumaraswamy inflacionada em zero e/ou um. Além de estudos de simulação, os autores apresentam uma aplicação a dados reais que retrata condições de saneamento (acessibilidade a água encanada e banheiro dentro da residência) para os municípios brasileiros. Para essa distribuição, os autores Bayer et al. (2021) propuseram o modelo de regressão Kumaraswamy inflacionado, com aplicação a dados de abastecimento de água e saneamento no Brasil. O modelo é usado para avaliar o impacto de diferentes variáveis condicionantes na proporção de pessoas que vivem em domicílios com abastecimento de água e esgoto inadequados nos municípios. Uma outra proposta é a distribuição Lindley unitária inflacionada em zero e um (BAPAT; BHARDWAJ, 2021). Os autores discutem algumas propriedades e apresentam estudos de simulação. Além disso, também comparam o ajuste da distribuição proposta com as distribuições beta e Kumaraswamy inflacionadas. Martinez-Florez et al. (2022a) apresentam a distribuição Birnbaum-Saunders unitária bimodal e fazem uma aplicação a dados inflacionados em zero e um. A partir de aplicações a dados reais e levando em conta a flexibilidade da distribuição Birnbaum-Saunders unitária bimodal, os autores evidenciaram um melhor desempenho em utilizar o modelo de regressão baseado na distribuição proposta em comparação ao modelo de regressão beta inflacionado. Ainda, Martinez-Florez et al. (2022b) propuseram a distribuição Birnbaum-Saunders unitária inflacionada e mostraram, através de uma aplicação, a vantagem de usar a distribuição proposta em relação à distribuição beta inflacionada. Liu et al. (2020b) propuseram a distribuição simplex inflacionada em zero e um para modelar proporções contínuas com excesso de zeros e/ou uns. Os autores também desenvolveram o modelo de regressão simplex inflacionado, apresentaram estudos de simulação e discutiram uma aplicação a dados de permanência hospitalar.

Embora tenham ocorrido avanços importantes com relação a modelagem de dados no intervalo unitário com inflação de zeros e/ou uns, ainda é possível ampliar os métodos para o estudo de variáveis dessa natureza. Neste artigo, propomos a distribuição gama unitária inflacionada em zero e/ou um, com uma abordagem diferente da apresentada nos trabalhos citados anteriormente. Consideramos uma reparametrização semelhante ao artigo de Bayes e Valdivieso (2016), em que a distribuição é expressa em termos da média da distribuição gama unitária inflacionada. Essa parametrização torna-se mais atrativa para fins de modelagem de regressão e gráficos de controle, por exemplo, uma vez que temos o interesse em modelar a média da distribuição.

O artigo está organizado como segue. Na Seção 2.2, apresentamos a distribuição gama unitária inflacionada e suas principais propriedades, dentre elas a função densidade de probabilidade, função de distribuição acumulada, valor esperado, variância e aspectos inferenciais. Na Seção 2.3, realizamos um extenso estudo de simulação de Monte Carlo para avaliar o comportamento dos estimadores pontuais. Adicionalmente, são apresentados resultados de simulação para avaliar a performance dos intervalos de confiança e testes de hipóteses assintóticos em amostras de tamanho finito. Na Seção 2.4, apresentamos a utilidade da distribuição proposta através de aplicações a conjuntos de dados reais. Por fim, na Seção 2.5, são feitas algumas considerações finais.

2.2 DISTRIBUIÇÃO GAMA UNITÁRIA

A distribuição gama unitária, proposta por Grassia (1977), é adequada para modelar variáveis no intervalo unitário (0,1). Mousa et al. (2016) propuseram uma reparametrização da distribuição gama unitária com densidade dada por:

$$gu(y; \mu, \phi) = \frac{\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}\right)^\phi}{\Gamma(\phi)} y^{\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}-1} \left[\log\left(\frac{1}{y}\right)\right]^{\phi-1}, \quad 0 < y, \mu < 1, \quad \text{e } \phi > 0. \quad (1)$$

O r -ésimo momento de uma variável aleatória que segue distribuição gama unitária é dado por:

$$E(Y^r) = \frac{\mu}{\left[\mu^{1/\phi} + (1 - \mu^{1/\phi})r\right]^\phi}, \quad (2)$$

com valor esperado e variância dados, respectivamente, por $E(Y) = \mu$ e $Var(Y) = \mu\{(1/(2-\mu^{1/\phi})^\phi) - \mu\}$. Sua função de distribuição acumulada pode ser escrita da seguinte forma:

$$GU(y; \mu, \phi) = 1 - \frac{\gamma\left(\phi, \frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}(-\log y)\right)}{\Gamma(\phi)}, \quad (3)$$

em que $\gamma(\cdot, \cdot)$ é a função gama incompleta inferior, definida por $\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$. A distribuição gama unitária é bastante flexível, podendo assumir diversas formas de acordo com os valores de seus parâmetros. A Figura 2.1 ilustra a densidade gama unitária para diferentes valores de μ e ϕ .

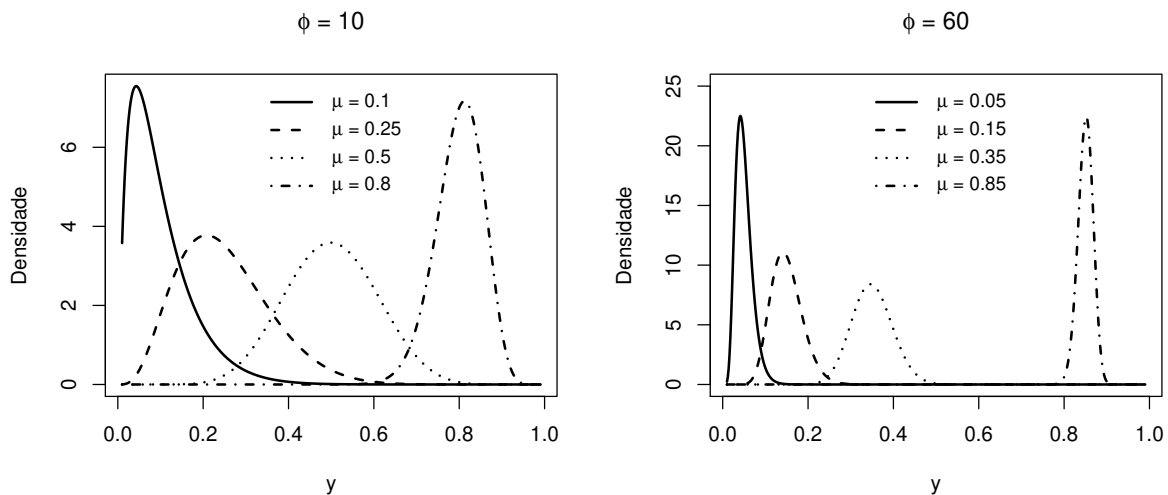


Figura 2.1 – Função densidade da gama unitária para diferentes valores de μ e ϕ .

Conforme exposto, a distribuição gama unitária é adequada para modelar variáveis que pertençam ao intervalo unitário, isto é, $y \in (0,1)$. Entretanto, em algumas situações o interesse recai em investigar variáveis que assumem valores nos intervalos $(0, 1]$, $[0, 1)$ ou $[0, 1]$ e, nestes casos, é recomendada a utilização de distribuições inflacionadas. Essa metodologia envolve a mistura entre duas

distribuições, sendo a distribuição de Bernoulli e uma distribuição contínua. Assim, fazendo uso dessas distribuições, é possível atribuir probabilidades aos pontos onde ocorre a inflação, podendo ser em zero, um ou ambos.

2.2.1 Distribuição proposta

A função densidade de probabilidade de uma variável aleatória Y que segue distribuição gama unitária inflacionada em zero e um, $Y \sim \text{GUI}_m(\delta_0, \delta_1, \mu, \phi)$, pode ser expressa por:

$$f(y; \delta_0, \delta_1, \mu, \phi) = \begin{cases} \delta_0, & \text{se } y = 0, \\ (1 - \delta_0 - \delta_1)gu(y; \mu; \phi), & \text{se } y \in (0, 1), \\ \delta_1, & \text{se } y = 1. \end{cases} \quad (4)$$

em que δ_0 denota a probabilidade de Y assumir valor 0, δ_1 a probabilidade de Y ser igual a 1, $gu(y; \mu; \phi)$ é a função densidade da distribuição gama unitária apresentada em (1), $0 < \delta_0 < 1$, $0 < \delta_1 < 1$, $0 < \delta_0 + \delta_1 < 1$, $\mu \in (0, 1)$, isto é, μ é a média condicional da variável resposta dado que $y \in (0, 1)$ e $\phi > 0$. A média e a variância de uma variável aleatória $Y \sim \text{GUI}_m(\delta_0, \delta_1, \mu, \phi)$ são expressas, respectivamente, por:

$$E(Y) = \delta_1 + (1 - \delta_0 - \delta_1)\mu$$

e

$$\text{Var}(Y) = \delta_1(1 - \delta_1) + (1 - \delta_0 - \delta_1) \left[\sigma_{GU}^2 - 2\mu\delta_1 + \mu^2(\delta_0 + \delta_1) \right],$$

em que σ_{GU}^2 é a variância de uma variável aleatória que segue distribuição gama unitária, isto é,

$$\sigma_{GU}^2 = \mu \left\{ \frac{1}{(2 - \mu^{1/\phi})^\phi} - \mu \right\}.$$

Usualmente, as distribuições inflacionadas têm como parâmetro a média condicional e não a média da distribuição inflacionada. No entanto, para modelos de regressão e gráficos de controle, por exemplo, torna-se mais interessante modelar a média da distribuição inflacionada, já que nos dois contextos estamos interessados em modelar ou monitorar a média do modelo e do processo, respectivamente. Se considerarmos um modelo de regressão no qual a variável resposta segue distribuição $\text{GUI}_m(\delta_0, \delta_1, \mu, \phi)$, o parâmetro μ representa a média condicional, ou seja, $\mu = E(Y|y \in (0, 1))$. Assim, considerando uma variável que pode assumir valores no intervalo $[0, 1]$, como é o caso do índice de esgoto tratado em um município, a média estimada sob essa abordagem não leva em consideração os municípios em que o índice for igual a zero ou igual 1 (100% de esgoto tratado).

Neste sentido, de forma similar ao trabalho de Bayes e Valdivieso (2016), reparametrizamos a distribuição gama unitária inflacionada (4) com o objetivo de modelar a média da distribuição gama

unitária inflacionada e não a média da distribuição gama unitária. Note que o valor esperado de $Y \sim \text{GUI}_m(\delta_0, \delta_1, \mu, \phi)$, $\gamma = E(Y)$, depende de μ e é restrito aos valores de δ_0 e δ_1 , de acordo com

$$\delta_1 < \gamma < 1 - \delta_0.$$

Assim, para permitir que os parâmetros do modelo sejam estimados sem restrições, sugerimos a seguinte parametrização

$$\gamma = \delta_1 + (1 - \delta_0 - \delta_1)\mu, \quad \alpha_0 = \frac{\delta_0}{1 - \gamma} \quad \text{e} \quad \alpha_1 = \frac{\delta_1}{\gamma}, \quad (5)$$

com $\gamma \in (0, 1)$, $\alpha_0 \in (0, 1)$, $\alpha_1 \in (0, 1)$ e $\phi > 0$. Considerando a nova parametrização e usando a notação $Y \sim \text{GUI}_m(\alpha_0, \alpha_1, \gamma, \phi)$, tem-se que

$$E(Y) = \gamma,$$

$$\text{Var}(Y) = \alpha_1\gamma(1 - \alpha_1\gamma) + \gamma(1 - \alpha_1) \left[\left(2 - \left(\frac{\gamma(1 - \alpha_1)}{1 - \alpha_0(1 - \gamma) - \alpha_1\gamma} \right)^{1/\phi} \right)^{-\phi} - \gamma(1 + \alpha_1) \right],$$

são, respectivamente, a média e a variância da distribuição gama unitária inflacionada reparametrizada. O r -ésimo momento de uma variável aleatória com distribuição gama unitária inflacionada é dado por

$$E(Y^r) = \alpha_1\gamma + (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)\mu_r,$$

em que μ_r é o r -ésimo momento ao redor de zero da distribuição $gu(\mu, \phi)$ dada em (1). Sua função densidade de probabilidade é dada por

$$f(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} \alpha_0(1 - \gamma), & \text{se } y = 0 \\ (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)gu\left(y; \frac{\gamma(1 - \alpha_1)}{1 - \alpha_0(1 - \gamma) - \alpha_1\gamma}, \phi\right), & \text{se } y \in (0, 1) \\ \alpha_1\gamma, & \text{se } y = 1. \end{cases} \quad (6)$$

Uma outra vantagem dessa proposta é que a expressão anterior pode ser estendida para os casos em que a variável de interesse assume apenas um dos extremos, conforme Equação (4), com $\delta_0 = 0$ e $\delta_1 > 0$ (distribuição gama unitária inflacionada em um), $\delta_0 > 0$ e $\delta_1 = 0$ (distribuição gama unitária inflacionada em zero) ou $\delta_0 = \delta_1 = 0$ (distribuição gama unitária). Quando a distribuição é inflacionada em um, temos que $\gamma = \delta_1 + (1 - \delta_1)\mu$ e $\alpha_1 = \frac{\delta_1}{\gamma}$, por outro lado, quando a distribuição é inflacionada em zero, temos que $\gamma = (1 - \delta_0)\mu$ e $\alpha_0 = \frac{\delta_0}{1 - \gamma}$. A função de distribuição acumulada de uma variável aleatória $Y \sim \text{GUI}_m(\alpha_0, \alpha_1, \gamma, \phi)$ é expressa por:

$$F(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} 0, & \text{se } y < 0 \\ \alpha_0(1 - \gamma), & \text{se } y = 0 \\ \alpha_0(1 - \gamma) + (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)GU\left(y; \frac{\gamma(1 - \alpha_1)}{1 - \alpha_0(1 - \gamma) - \alpha_1\gamma}, \phi\right), & \text{se } 0 < y < 1 \\ 1, & \text{se } y \geq 1. \end{cases}$$

A Figura 2.2 apresenta diferentes densidades da variável aleatória que segue distribuição GUI_m em zero e/ou um para diferentes valores de α_0 , α_1 , γ e ϕ .

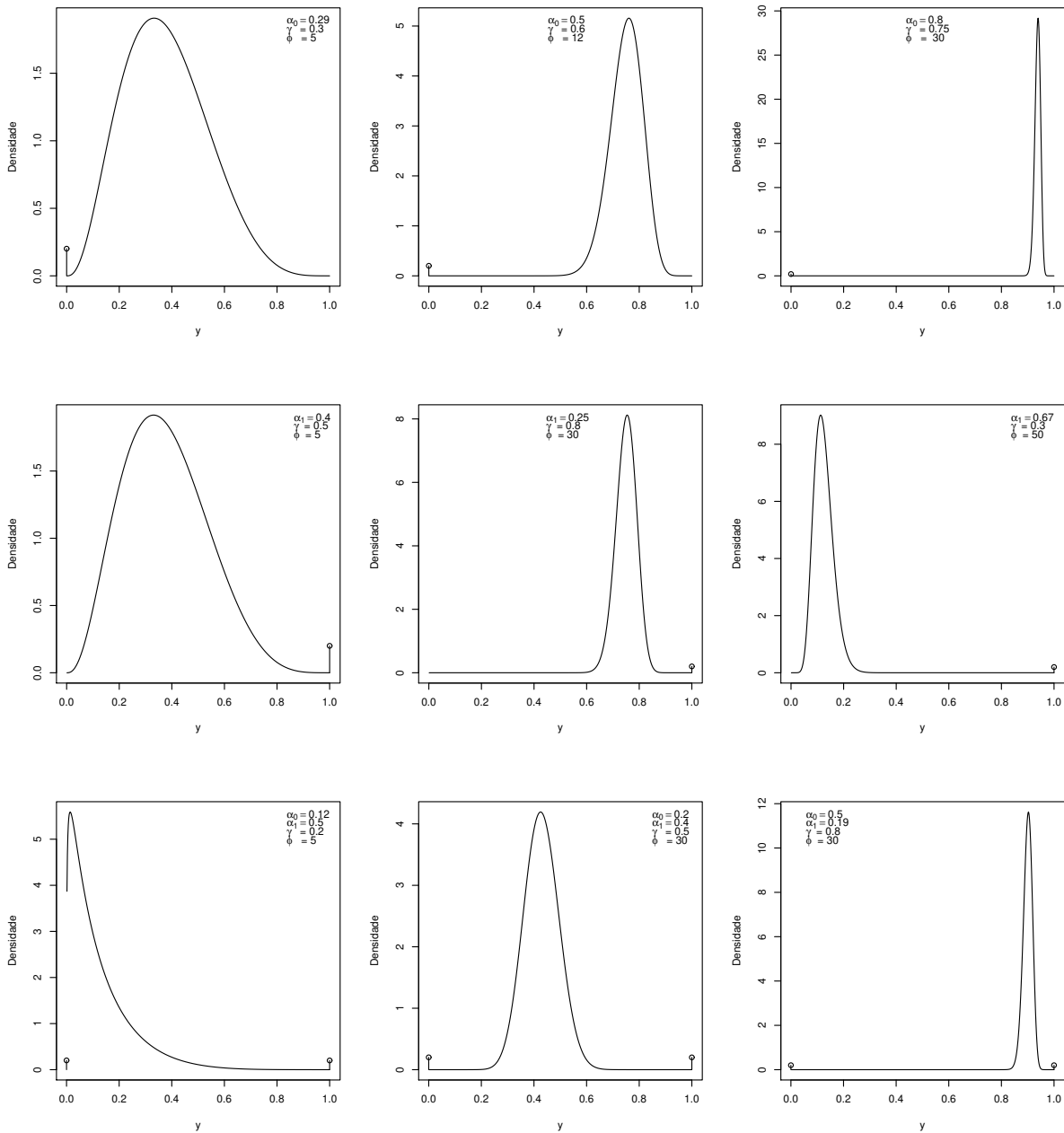


Figura 2.2 – Função densidade de probabilidade da gama unitária inflacionada para diferentes valores de α_0 , α_1 , γ e ϕ .

2.2.2 Estimação

Nesta seção, apresentamos o procedimento para estimação dos parâmetros que indexam a distribuição gama unitária inflacionada reparametrizada. A estimação do vetor de parâmetros $\theta = (\alpha_0, \alpha_1, \gamma, \phi)^\top$ é realizada através do método de máxima verossimilhança. Dessa forma, podemos reescrever a função densidade de probabilidade de Y como

$$\begin{aligned} f(y; \alpha_0, \alpha_1, \gamma, \phi) &= [\alpha_0(1-\gamma)]^{\mathbb{1}_{\{0\}}(y)} [\alpha_1\gamma]^{\mathbb{1}_{\{1\}}(y)} \\ &\times [(1-\alpha_0(1-\gamma) - \alpha_1\gamma)gu(y; \check{\mu}, \phi)]^{(1-\mathbb{1}_{\{0\}}(y))(1-\mathbb{1}_{\{1\}}(y))}, \end{aligned} \quad (7)$$

em que $\mathbb{1}_A(y)$ é a função indicadora que é igual a 1 se $y \in A$ e 0 se $y \notin A$ e $\check{\mu} = \frac{\gamma(1-\alpha_1)}{1-\alpha_0(1-\gamma)-\alpha_1\gamma}$.

Seja uma amostra aleatória $\mathbf{y} = (y_1, \dots, y_n)^\top$. A função de verossimilhança, $L(\theta)$, de $\theta = (\alpha_0, \alpha_1, \gamma, \phi)^\top$ dada a amostra é expressa por:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(y_i; \alpha_0, \alpha_1, \gamma, \phi) \\ &= [\alpha_0(1-\gamma)]^{\sum_{i=1}^n \mathbb{1}_{\{0\}}(y_i)} [\alpha_1\gamma]^{\sum_{i=1}^n \mathbb{1}_{\{1\}}(y_i)} \\ &\times \prod_{i=1}^n [(1-\alpha_0(1-\gamma) - \alpha_1\gamma)gu(y_i; \check{\mu}, \phi)]^{(1-\mathbb{1}_{\{0\}}(y_i))(1-\mathbb{1}_{\{1\}}(y_i))}. \end{aligned}$$

Assim, podemos escrever a função de log-verossimilhança, $\ell(\theta)$, como:

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n \mathbb{1}_{\{0\}}(y_i) \log [\alpha_0(1-\gamma)] + \sum_{i=1}^n \mathbb{1}_{\{1\}}(y_i) \log [\alpha_1\gamma] \\ &+ \sum_{i=1}^n (1-\mathbb{1}_{\{0\}}(y_i))(1-\mathbb{1}_{\{1\}}(y_i)) \log c + \sum_{i=1}^n (1-\mathbb{1}_{\{0\}}(y_i))(1-\mathbb{1}_{\{1\}}(y_i)) \log gu(y_i; \check{\mu}, \phi), \end{aligned}$$

em que $c = 1 - \alpha_0(1-\gamma) - \alpha_1\gamma$ e

$$\log gu(y_i; \check{\mu}, \phi) = \phi \log \left(\frac{\check{\mu}^{1/\phi}}{1 - \check{\mu}^{1/\phi}} \right) - \log \Gamma(\phi) + \log y_i \left(\frac{\check{\mu}^{1/\phi}}{1 - \check{\mu}^{1/\phi}} - 1 \right) + (\phi - 1) \log \left[\log \left(\frac{1}{y_i} \right) \right]$$

é a log-densidade da distribuição gama unitária. Note que

$$\frac{\partial \log gu(y_i; \check{\mu}, \phi)}{\partial \check{\mu}} = \frac{d(1+d)}{\check{\mu}\phi} (y_i^* - \mu^*), \quad (8)$$

em que $y^* = \log y_i$, $\mu^* = -\frac{\phi}{d}$ e $d = \frac{\check{\mu}^{1/\phi}}{1 - \check{\mu}^{1/\phi}}$. A função escore, que é obtida pela diferenciação da função de log-verossimilhança com relação a cada parâmetro, é denotada por $U(\theta) = \left(\frac{\partial \ell(\theta)}{\partial \alpha_0}, \frac{\partial \ell(\theta)}{\partial \alpha_1}, \frac{\partial \ell(\theta)}{\partial \gamma}, \frac{\partial \ell(\theta)}{\partial \phi} \right)^\top$.

Os estimadores de máxima verossimilhança são obtidos como solução de $U(\theta) = 0$. No entanto, como os estimadores de máxima verossimilhança não podem ser expressos em forma fechada, eles podem ser obtidos numericamente maximizando a função de log-verossimilhança usando um método de otimização não linear. O método quase-Newton BFGS é comumente usado para maximizar funções de log-verossimilhança, para mais detalhes consultar Jorge e Stephen (2006) e Press (2007). As expressões para as primeiras derivadas com relação a cada um dos parâmetros são apresentadas no Apêndice A.

Seja $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\gamma}, \hat{\phi})^\top$ o estimador de máxima verossimilhança de θ . Em grandes amostras e sob certas condições de regularidade (SEN; SINGER, 1994), a distribuição de $\hat{\theta}$ converge em distribuição para uma Normal, isto é, $\hat{\theta} \stackrel{a}{\sim} N(\theta, K(\theta)^{-1})$, em que $K(\theta)$ é a matriz dada em (9) e $\stackrel{a}{\sim}$ é a distribuição assintótica. Usando o resultado anterior é possível construir intervalos de confiança baseados na normalidade assintótica. Por exemplo, temos que $\hat{\gamma} \pm z_{1-\frac{\delta}{2}}(\widehat{K}^{\gamma\gamma})^{1/2}$ são os limites de confiança assintóticos para γ de coeficiente de confiança $(1 - \delta) \times 100\%$, sendo $z_{1-\frac{\delta}{2}}$ o quantil $(1 - \delta/2)$ da distribuição $N(0, 1)$ e $\widehat{K}^{\gamma\gamma}$ a variância assintótica de $\hat{\gamma}$ obtida do inverso da matriz $K(\theta)$ dada em (9) sendo avaliada no estimador de máxima verossimilhança. A matriz de informação de Fisher para a distribuição gama unitária inflacionada reparametrizada é

$$K(\theta) = \begin{pmatrix} K_{\gamma\gamma} & K_{\gamma\alpha_0} & K_{\gamma\alpha_1} & K_{\gamma\phi} \\ K_{\alpha_0\gamma} & K_{\alpha_0\alpha_0} & K_{\alpha_0\alpha_1} & K_{\alpha_0\phi} \\ K_{\alpha_1\gamma} & K_{\alpha_1\alpha_0} & K_{\alpha_1\alpha_1} & K_{\alpha_1\phi} \\ K_{\phi\gamma} & K_{\phi\alpha_0} & K_{\phi\alpha_1} & K_{\phi\phi} \end{pmatrix}. \quad (9)$$

As expressões para os elementos da matriz podem ser encontradas no Apêndice A.

2.2.3 Testes de hipóteses

A normalidade assintótica de $\hat{\theta}$ também pode ser usada para a construção de testes de hipóteses. No contexto de testes de hipóteses o objetivo é utilizar um conjunto de dados observados para fazer afirmações a respeito de sua distribuição ou de seu parâmetro θ (LEHMANN; ROMANO, 1986). Para a construção de um teste de hipóteses é necessário conhecermos a distribuição da estatística de teste sob a hipótese nula; quando essa distribuição exata é analiticamente difícil de ser obtida ou até mesmo impossível de ser definida, fazemos uso de distribuições aproximadas a partir da teoria assintótica. Os testes assintóticos mais conhecidos na literatura são os da razão de verossimilhanças (RV), Wald (W), escore (S_R) e gradiente (S_G). O teste gradiente, apesar de mais recente que os demais (TERRELL, 2002), tem sido utilizado em diversos modelos na literatura devido sua facilidade de implementação. Mais detalhes sobre os testes podem ser encontrados em Buse (1982), Cox e Hinkley (1979), Welsh (2011) e Lemonte (2016).

Considere que o interesse esteja em fazer inferência em um conjunto de parâmetros. Assim, seja $\theta = (\theta_1^\top, \theta_2^\top)^\top$, em que θ_1 é um vetor $r \times 1$ de parâmetros de interesse e θ_2 é um vetor $(m - r) \times 1$

de parâmetros de perturbação. Suponha que desejamos testar a hipótese nula $\mathcal{H}_0 : \theta_1 = \theta_1^{(0)}$ versus a hipótese alternativa $\mathcal{H}_1 : \theta_1 \neq \theta_1^{(0)}$. Seja $\hat{\theta}$ o estimador de máxima verossimilhança irrestrito de θ e $\tilde{\theta} = \left(\theta_1^{(0)\top}, \tilde{\theta}_2^\top \right)^\top$ o estimador de máxima verossimilhança restrito de θ . O estimador restrito é obtido impondo a hipótese nula (\mathcal{H}_0). Dessa forma, as estatísticas da razão de verossimilhanças e Wald são dadas, respectivamente, por

$$RV = 2 \left[\ell(\hat{\theta}) - \ell(\tilde{\theta}) \right],$$

e

$$W = \left(\hat{\theta}_1 - \theta_1^{(0)} \right)^\top \left[K^{rr}(\hat{\theta}) \right]^{-1} \left(\hat{\theta}_1 - \theta_1^{(0)} \right).$$

A estatística para o teste escore de Rao pode ser expressa por

$$S_R = U_r(\tilde{\theta})^\top K^{rr}(\tilde{\theta}) U_r(\tilde{\theta}),$$

em que $K^{rr}(\hat{\theta})$ é o bloco $r \times r$ da inversa da matriz de informação de Fisher avaliada em $\hat{\theta}$, $U_r(\tilde{\theta})$ e $K^{rr}(\tilde{\theta})$ denotam o vetor $r \times 1$ que contém r elementos da função escore e o bloco $r \times r$ da inversa da matriz de informação de Fisher, respectivamente, avaliada em $\tilde{\theta}$.

Por fim, temos a estatística gradiente

$$S_G = U(\theta_0)^\top (\hat{\theta} - \theta_0).$$

Sob a hipótese nula e considerando condições de regularidade usuais, as estatísticas RV , W , S_R e S_G convergem em distribuição para uma distribuição qui-quadrado com r graus de liberdades. Assim, os testes apresentados podem ser realizados de acordo com os valores críticos assintóticos obtidos da distribuição χ_r^2 .

2.3 ESTUDO DE SIMULAÇÃO

Nesta seção, são apresentados resultados de simulações de Monte Carlo para avaliar o comportamento dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição proposta. Também são conduzidas simulações para avaliar a performance de intervalos de confiança assintóticos e dos testes de hipóteses Wald, razão de verossimilhanças, escore e gradiente. A maximização da função de log-verossimilhança foi realizada usando o método quasi-Newton BFGS com derivadas analíticas, usualmente considerado como método de melhor desempenho (MITTELHAMMER et al., 2000). Os resultados foram baseados em 10.000 réplicas de Monte Carlo. Os cenários

foram definidos considerando a distribuição gama unitária inflacionada em um (GUI_{m1}) e a distribuição gama unitária inflacionada em zero e um (GUI_{m01}).

Os valores iniciais para os parâmetros γ , α_0 e α_1 foram, respectivamente, \bar{y} , $\alpha_0 = \delta_0/(1 - \bar{y})$, $\alpha_1 = \delta_1/\bar{y}$, em que δ_0 e δ_1 representam as proporções de zeros e uns na amostra, respectivamente, e \bar{y} é a média amostral. Para o parâmetro ϕ o chute inicial foi dado baseando-se no artigo de Rocha et al. (2021). Na proposta dos autores foram usadas propriedades da família exponencial gama unitária para atribuir um valor inicial para ϕ .

Foram considerados os seguintes tamanhos de amostra: $n = 40, 80, 120$ e 200 . Inicialmente, nos concentramos na avaliação das estimativas pontuais calculando a média, o viés relativo e o erro quadrático médio dos estimadores de máxima verossimilhança de γ , α_0 , α_1 e ϕ . O erro quadrático médio é calculado como a diferença quadrática média entre o valor estimado e o valor real do parâmetro, enquanto o viés relativo é calculado como a diferença entre a estimativa média e o verdadeiro valor do parâmetro dividido pelo último. Na sequência, foram avaliadas as taxas de coberturas de intervalos de confiança e taxas de não coberturas. Foram consideradas probabilidades nominais de cobertura iguais a 90% e 95%. Por último, discutimos os resultados para os testes de hipóteses assintóticos com taxas de rejeição da hipótese nula iguais a 1%, 5% e 10%. Os resultados foram obtidos usando a linguagem de programação R (R Development Core Team, 2021).

2.3.1 Resultados numéricos

Na Tabela 2.1 encontram-se os resultados do estudo de simulação para avaliar o comportamento dos estimadores pontuais da distribuição GUI_{m1} , considerando os percentuais de uns na amostra iguais a 8% e 20%. Nesse cenário, Tabela 2.1, com média $\gamma = 0,8$ e $\phi = 30$, verificamos que as médias dos estimadores se aproximaram dos valores reais à medida que o tamanho da amostra aumenta. Comparativamente, temos o cenário com média $\gamma = 0,8$ e parâmetro de precisão $\phi = 12$. De forma similar, em geral, observamos que as estimativas dos parâmetros ficam próximas dos valores reais. Assim, não verificamos diferenças expressivas dos resultados encontrados no que se refere a redução da precisão.

A avaliação dos estimadores para a distribuição GUI_{m01} são apresentados na Tabela 2.2. Os percentuais de zeros e uns na amostra somaram 10% e 25%. É possível notar que as médias dos estimadores se aproximam dos valores reais, mesmo quando $n = 40$; e tornam-se ainda mais precisas à medida que o tamanho da amostra aumenta (Tabela 2.2). Além disso, é possível verificar que o aumento no parâmetro de precisão não causou mudanças expressivas nas estimativas dos parâmetros.

Tabela 2.1 – Média dos estimadores de máxima verossimilhança que indexam a distribuição GUI_{m1} ; $\gamma = 0,8$, $\phi = 12$ e $\phi = 30$.

ϕ	α_1	Estimador	n			
			40	80	120	200
12	0,10 ($\delta_1 = 8\%$)	$\hat{\gamma}$	0,7999	0,8001	0,8000	0,8000
		$\hat{\alpha}_1$	0,0990	0,0996	0,0999	0,0996
		$\hat{\phi}$	13,0752	12,4892	12,3308	12,1949
	0,25 ($\delta_1 = 20\%$)	$\hat{\gamma}$	0,7999	0,8002	0,8000	0,7999
		$\hat{\alpha}_1$	0,2484	0,2500	0,2496	0,2494
		$\hat{\phi}$	13,2772	12,6127	12,3838	12,2252
30	0,10 ($\delta_1 = 8\%$)	$\hat{\gamma}$	0,7999	0,8001	0,8000	0,8000
		$\hat{\alpha}_1$	0,0990	0,0996	0,0999	0,0996
		$\hat{\phi}$	32,7117	31,2332	30,8348	30,4913
	0,25 ($\delta_1 = 20\%$)	$\hat{\gamma}$	0,7999	0,8002	0,8000	0,7999
		$\hat{\alpha}_1$	0,2484	0,2500	0,2496	0,2494
		$\hat{\phi}$	33,2206	31,5377	30,9675	30,5695

Tabela 2.2 – Média dos estimadores de máxima verossimilhança que indexam a distribuição GUI_{m01} ; $\gamma = 0,5$, $\phi = 12$ e $\phi = 30$.

ϕ	α	Estimador	n			
			40	80	120	200
12	$\alpha_0 = 0,10$; $\alpha_1 = 0,10$ ($\delta_0 = 5\%$; $\delta_1 = 5\%$)	$\hat{\gamma}$	0,4998	0,5003	0,5000	0,4999
		$\hat{\alpha}_0$	0,0977	0,0985	0,0998	0,0997
		$\hat{\alpha}_1$	0,0975	0,0990	0,0994	0,0992
	$\alpha_0 = 0,26$; $\alpha_1 = 0,24$ ($\delta_0 = 13\%$; $\delta_1 = 12\%$)	$\hat{\phi}$	13,0941	12,5175	12,3506	12,1885
		$\hat{\gamma}$	0,4998	0,5005	0,5000	0,4998
		$\hat{\alpha}_0$	0,2546	0,2565	0,2583	0,2593
30	$\alpha_0 = 0,10$; $\alpha_1 = 0,10$ ($\delta_0 = 5\%$; $\delta_1 = 5\%$)	$\hat{\alpha}_1$	0,2348	0,2382	0,2386	0,2386
		$\hat{\phi}$	13,3033	12,5978	12,3917	12,2388
		$\hat{\gamma}$	0,4999	0,5003	0,4999	0,4999
	$\alpha_0 = 0,24$; $\alpha_1 = 0,26$ ($\delta_0 = 12\%$; $\delta_1 = 13\%$)	$\hat{\alpha}_0$	0,0976	0,0984	0,0998	0,0997
		$\hat{\alpha}_1$	0,0975	0,0990	0,0994	0,0992
		$\hat{\phi}$	32,7565	31,2955	30,8874	30,4700
	$\hat{\gamma}$	0,4999	0,5005	0,5000	0,4998	
	$\hat{\alpha}_0$	0,2350	0,2367	0,2384	0,2395	
	$\hat{\alpha}_1$	0,2549	0,2580	0,2584	0,2587	
		$\hat{\phi}$	33,3244	31,5537	31,0086	30,6282

Nas Figuras 2.3 e 2.4 podemos observar o comportamento do EQM e viés relativo (%) dos estimadores considerando a distribuição GUI_{m1} . É possível observar que o EQM e viés relativo de $\hat{\gamma}$ e $\hat{\alpha}_1$ foram baixos e para $\hat{\phi}$, os valores decrescem à medida que o tamanho da amostra aumenta. De modo geral, não foram observadas mudanças expressivas no comportamento dessas medidas quando temos uma proporção de uns maior na amostra ou uma redução no parâmetro de precisão (Figuras 2.3 e 2.4).

As Figuras 2.5 e 2.6 mostram os resultados para os cenários considerando a distribuição

GUI_{m01} . Comportamento semelhante ao descrito anteriormente pode ser observado nas Figuras 2.5 e 2.6, quando temos inflação de zeros e uns. Em geral, o EQM e o viés relativo dos estimadores tendem a zero à medida que o tamanho da amostra aumenta.

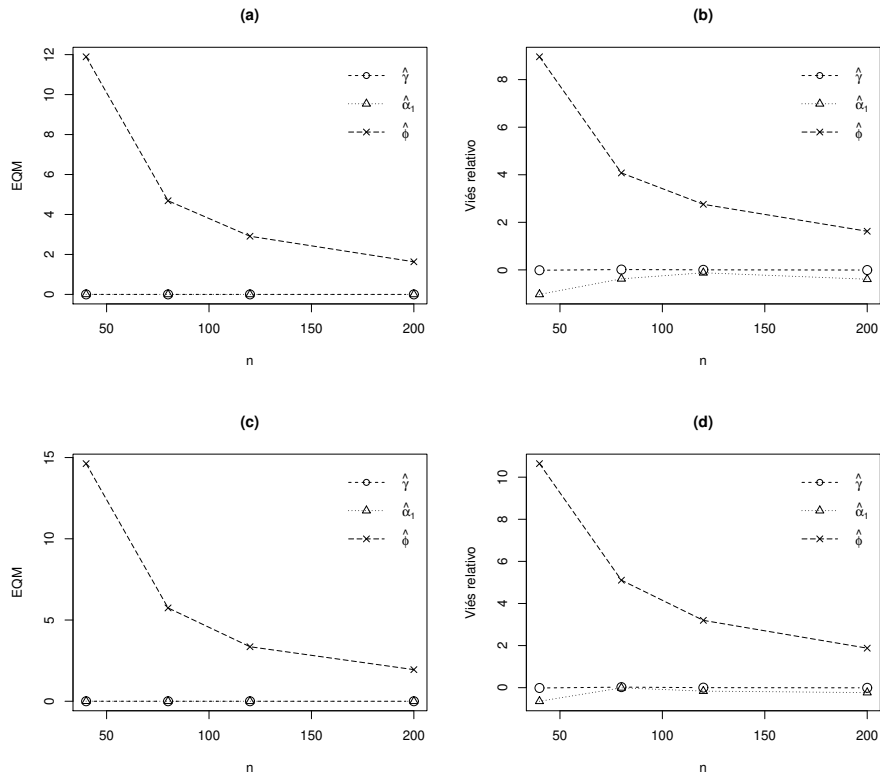


Figura 2.3 – Gráficos das medidas EQM e Viés relativo (%) quando $\gamma = 0,8$ e $\phi = 12$: (a) e (b) com $\alpha_1 = 0,1$; (c) e (d) com $\alpha_1 = 0,25$.

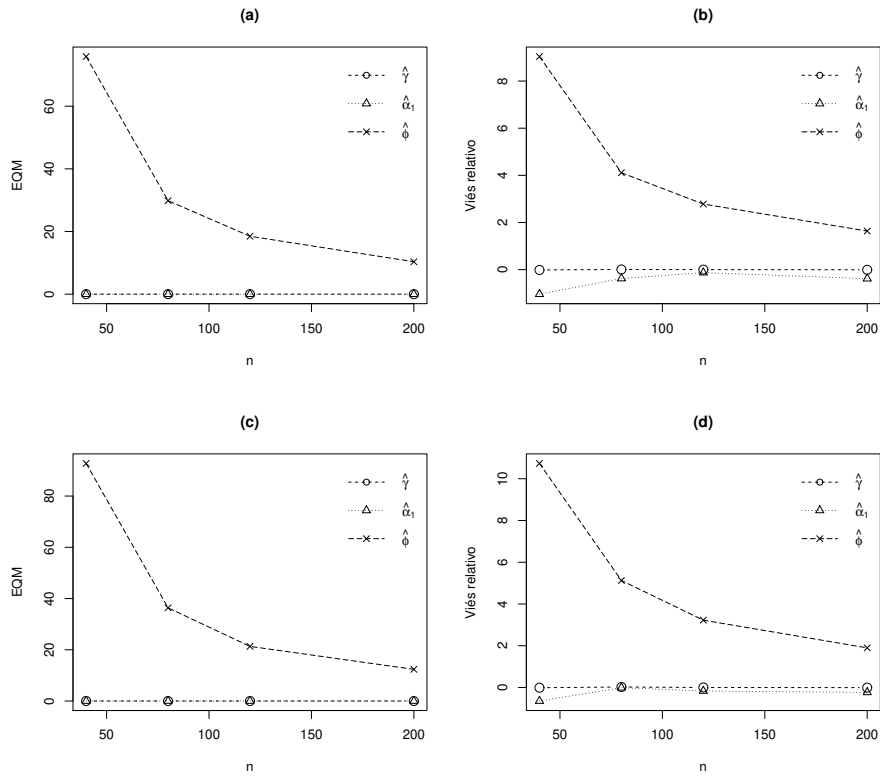


Figura 2.4 – Gráficos das medidas EQM e Viés relativo (%) quando $\gamma = 0,8$ e $\phi = 30$: (a) e (b) com $\alpha_1 = 0,1$; (c) e (d) $\alpha_1 = 0,25$.

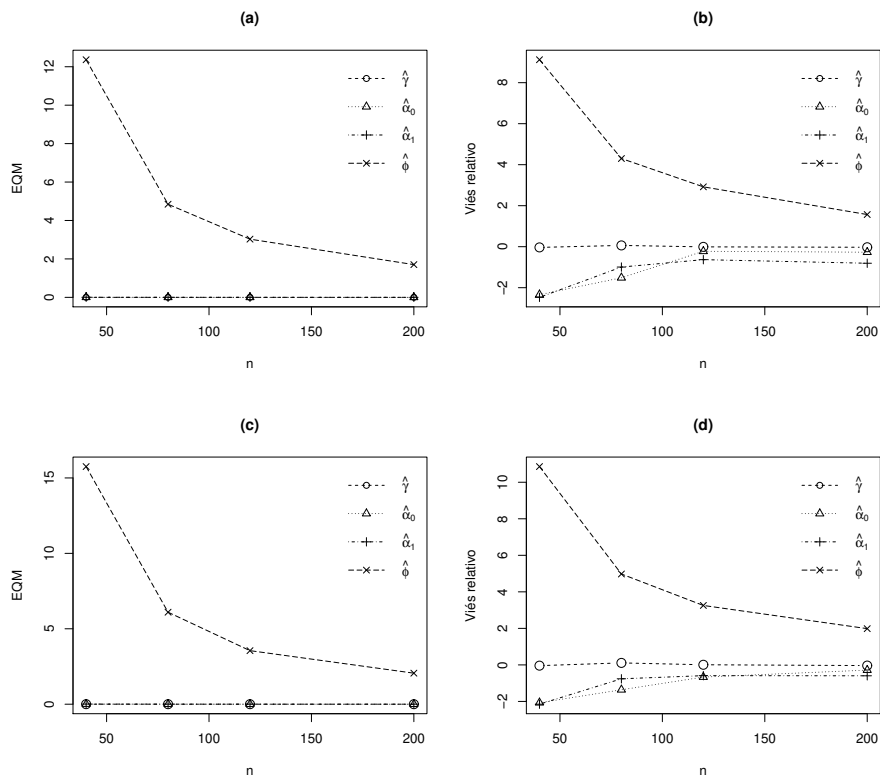


Figura 2.5 – Gráficos das medidas EQM e Viés relativo (%) quando $\gamma = 0,5$ e $\phi = 12$: (a) e (b) com $\alpha_0 = 0,1$ e $\alpha_1 = 0,1$; (c) e (d) com $\alpha_0 = 0,26$ e $\alpha_1 = 0,24$.

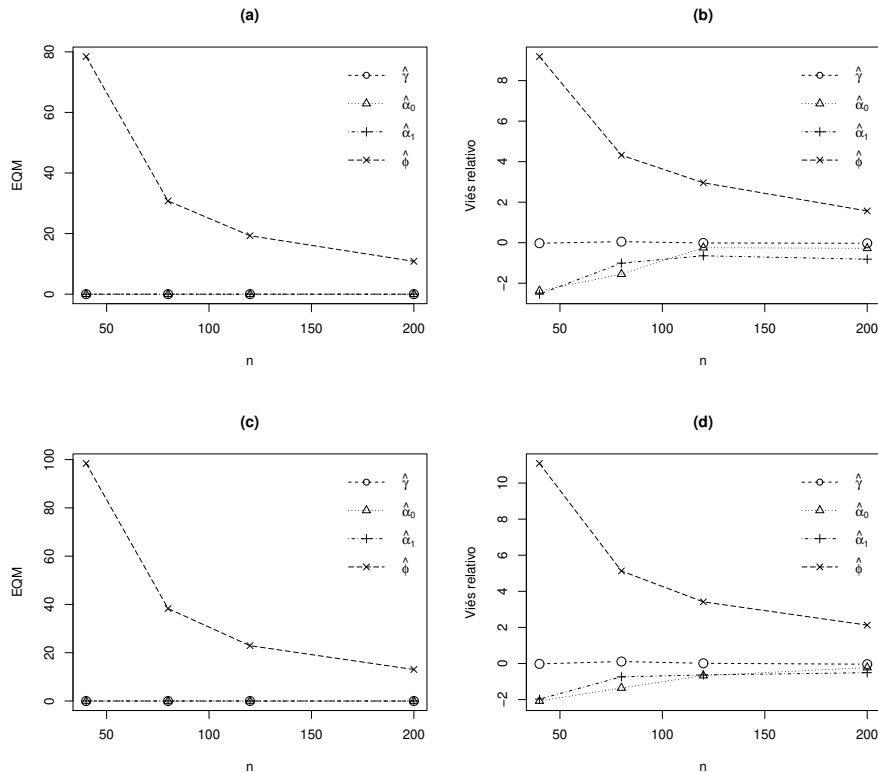


Figura 2.6 – Gráficos das medidas EQM e Viés relativo (%) quando $\gamma = 0,5$ e $\phi = 30$: (a) e (b) com $\alpha_0 = 0,1$ e $\alpha_1 = 0,1$; (c) e (d) com $\alpha_0 = 0,24$ e $\alpha_1 = 0,26$.

As Tabelas 2.3 e 2.4 apresentam as taxas empíricas de cobertura e não cobertura dos intervalos de confiança para os parâmetros da distribuição GUI_{m1} . Na Tabela 2.3, com $\phi = 12$ e $\delta_1 = 8\%$, é possível observar que as coberturas empíricas se aproximam das nominais com o aumento do tamanho de n . Por exemplo, considerando o parâmetro α_1 e nível de confiança de 95%, para $n = 40$ temos uma taxa de cobertura igual a 86,09% enquanto para $n = 200$ essa taxa é de 94,53%. Com base na Tabela 2.4, é possível notar que as taxas de coberturas empíricas ficam próximas das taxas nominais à medida que o tamanho da amostra aumenta. Inclusive para $n = 40$, os resultados já são satisfatórios, em especial, para os parâmetros γ e ϕ . As taxas de não cobertura também se tornam mais equilibradas à medida que o tamanho da amostra aumenta.

Com relação às taxas de cobertura de intervalos de confiança para a distribuição GUI_{m01} , visualizamos que o aumento na proporção de zeros e uns na amostra proporcionou taxas de cobertura relativamente maiores quando consideramos $n = 200$ (Tabela 2.5). Por exemplo, para $\delta_0 = \delta_1 = 5\%$ e $n = 200$, as taxas de cobertura para α_0 e α_1 foram 92,70% e 92,86%, respectivamente. Em contrapartida, quando $\delta_0 = 13\%$, $\delta_1 = 12\%$ e $n = 200$ (Tabela 2.6), todas as taxas de cobertura foram acima de 94%, considerando o nível de confiança de 95%. Se considerarmos outro tamanho de amostra, por exemplo, $n = 120$ e nível de confiança de 90%, também observamos que as taxas de cobertura para os parâmetros γ , α_0 e α_1 ficaram mais próximas dos níveis nominais considerados para uma maior proporção de zeros e uns, sendo 89,66%, 88,71%, 88,86% (Tabela 2.6), respectivamente, em relação aos valores 89,57%, 84,93% e 85,21% (Tabela 2.5). As maiores taxas de não cobertura foram obser-

vadas para os parâmetros α_0 e α_1 , com 7,71% (para direita) e 7,13% (para direita), respectivamente, considerando $n = 80$ e nível de confiança de 95% (Tabela 2.5). Para o nível de confiança de 90% e $n = 120$, ainda na Tabela 2.5, α_0 e α_1 apresentaram taxas de não cobertura à direita iguais a 12,52% e 12,18%, respectivamente. Com relação aos parâmetros γ e ϕ , observamos que as taxas de não cobertura foram relativamente baixas, não ultrapassando 3% para o nível de confiança de 95% e 6% para o nível de confiança de 90%.

Considerando $\gamma = 0,5$ e o aumento no parâmetro de precisão ($\phi = 30$), Tabela 2.7, verificamos que, de modo geral, as probabilidades empíricas de cobertura de todos os intervalos se aproximam de 95% quando o tamanho da amostra aumenta. Com um nível de confiança de 95%, percebe-se que as maiores taxas de não cobertura ocorrem para os parâmetros α_0 e α_1 , sendo especificamente taxas de não cobertura à direita. Esse comportamento também pode ser observado quando consideramos o nível de 90%. Por exemplo, para $n = 40$ e $n = 80$ as taxas de não cobertura à direita para o parâmetro α_0 foram, respectivamente, 11,16% e 8,29%.

Tabela 2.3 – Taxas de cobertura (%) empíricas e de não cobertura (para esquerda; para direita) dos intervalos de confiança, a níveis de confiança de 90% e 95%, dos parâmetros da distribuição $GUI_{m,1}$ com $\gamma = 0,8$ e $\phi = 12$.

α_1	$1 - \delta$	Parâmetro	n			
			40	80	120	200
0,1 ($\delta_1 = 8\%$)	95%	γ	95,07 (2,04; 2,88)	94,22 (2,12; 3,65)	94,69 (1,94; 3,36)	94,72 (1,99; 3,28)
		α_1	86,09 (0,70; 13,20)	89,73 (1,17; 9,09)	91,32 (1,22; 7,45)	94,53 (1,23; 4,23)
		ϕ	95,69 (1,63; 2,67)	95,28 (1,80; 2,91)	94,97 (2,16; 2,86)	95,30 (1,88; 2,81)
	90%	γ	89,96 (4,48; 5,55)	89,09 (4,38; 6,52)	89,33 (4,54; 6,12)	89,72 (4,42; 5,85)
		α_1	85,34 (1,45; 13,20)	86,44 (2,67; 10,88)	89,45 (3,09; 7,45)	89,92 (3,13; 6,94)
		ϕ	90,65 (4,54; 4,80)	90,16 (4,43; 5,40)	89,89 (4,84; 5,26)	90,60 (4,33; 5,06)

Tabela 2.4 – Taxas de cobertura (%) empíricas e de não cobertura (para esquerda; para direita) dos intervalos de confiança, a níveis de confiança de 90% e 95%, dos parâmetros da distribuição $GUI_{m,1}$ com $\gamma = 0,8$ e $\phi = 30$.

α_1	$1 - \delta$	Parâmetro	n			
			40	80	120	200
0,1 ($\delta_1 = 8\%$)	95%	γ	94,75 (1,32; 3,92)	93,54 (1,48; 4,97)	93,97 (1,59; 4,43)	94,49 (1,56; 3,94)
		α_1	86,07 (0,72; 13,2)	89,00 (1,17; 9,82)	91,23 (1,31; 7,45)	94,79 (1,16; 4,04)
		ϕ	95,75 (1,65; 2,59)	95,27 (1,81 ; 2,91)	95,08 (2,1; 2,81)	95,39 (1,82; 2,78)
		γ	89,9 (3,17; 6,92)	88,42 (3,47; 8,1)	89,05 (3,74; 7,2)	89,49 (3,99; 6,51)
		α_1	85,53 (1,26; 13,2)	86,44 (2,67; 10,88)	89,45 (3,09; 7,45)	89,92 (3,13; 6,94)
		ϕ	90,63 (4,53; 4,83)	90,11 (4,47; 5,41)	89,91 (4,93; 5,15)	90,55 (4,36; 5,08)
	90%	γ	93,21 (1,85; 4,93)	93,87 (2,07; 4,05)	94,44 (1,85; 3,70)	94,45 (2,19; 3,35)
		α_1	90,64 (1,97; 7,38)	93,05 (2,25; 4,69)	94,24 (1,98; 3,77)	94,27 (2,38; 3,34)
		ϕ	95,49 (1,57; 2,93)	95,04 (2,04; 2,91)	94,95 (2,24; 2,80)	95,00 (2,25; 2,74)
		γ	88,04 (3,98; 7,97)	89,38 (4,07; 6,54)	89,19 (4,37; 6,43)	88,99 (4,81; 6,19)
		α_1	88,20 (4,38; 7,41)	90,61 (3,97; 5,41)	88,88 (4,76; 6,35)	88,54 (5,12; 6,33)
		ϕ	90,70 (4,37; 4,92)	89,84 (4,98; 5,17)	89,75 (5,07; 5,17)	90,52 (4,78; 4,69)

Tabela 2.5 – Taxas de cobertura (%) empíricas e de não cobertura (para esquerda; para direita) dos intervalos de confiança, a níveis de confiança de 90% e 95%, dos parâmetros da distribuição GUI_{m01} com $\gamma = 0,5$ e $\phi = 12$.

α	$1 - \delta$	Parâmetro	n			
			40	80	120	200
$\alpha_0 = \alpha_1 = 0,10$ ($\delta_0 = \delta_1 = 5\%$)	95%	γ	97,95 (0,90; 1,14)	95,16 (2,42; 2,41)	94,96 (2,61; 2,42)	94,69 (2,71; 2,59)
		α_0	99,17 (0,82; 0,00)	91,50 (0,78; 7,71)	93,35 (0,88; 5,76)	92,70 (1,04; 6,25)
		α_1	99,15 (0,84; 0,00)	92,14 (0,72; 7,13)	93,52 (0,79; 5,68)	92,86 (0,94; 6,19)
		ϕ	95,46 (1,74; 2,79)	95,25 (1,98; 2,76)	95,20 (2,06; 2,73)	95,15 (2,14; 2,70)
		γ	94,37 (2,67; 2,95)	90,11 (4,91; 4,97)	89,57 (5,20; 5,22)	89,99 (5,03; 4,97)
		α_0	98,34 (1,65; 0,00)	90,22 (2,04; 7,73)	84,93 (2,54; 12,52)	88,59 (2,86; 8,54)
		α_1	98,42 (1,57; 0,00)	90,64 (2,20; 7,15)	85,21 (2,60; 12,18)	88,77 (2,62; 8,60)
		ϕ	90,55 (4,46; 4,98)	90,41 (4,77; 4,81)	90,07 (4,91; 5,01)	90,01 (5,03; 4,95)

Tabela 2.6 – Taxas de cobertura (%) empíricas e de não cobertura (para esquerda; para direita) dos intervalos de confiança, a níveis de confiança de 90% e 95%, dos parâmetros da distribuição GUI_{m01} com $\gamma = 0,5$ e $\phi = 12$.

α	$1 - \delta$	Parâmetro	n			
			40	80	120	200
$\alpha_0 = 0,26; \alpha_1 = 0,24$ ($\delta_0 = 13\%; \delta_1 = 12\%$)	95%	γ	94,36 (3,00; 2,63)	94,17 (3,09; 2,73)	94,65 (2,74; 2,6)	94,86 (2,58; 2,55)
		α_0	90,82 (1,87; 7,3)	93,06 (1,95; 4,98)	93,54 (2,09; 4,36)	94,10 (2,05; 3,84)
		α_1	91,80 (1,85; 6,34)	92,90 (1,89; 5,2)	94,00 (1,98; 4,01)	94,19 (2,04; 3,76)
		ϕ	95,70 (1,36; 2,93)	95,45 (1,68; 2,86)	94,79 (2,1; 3,1)	94,75 (2,15; 3,09)
		γ	88,79 (5,88; 5,32)	89,06 (5,73; 5,2)	89,66 (5,3; 5,03)	89,62 (5,34; 5,03)
		α_0	87,02 (3,86; 9,11)	87,95 (3,89; 8,15)	88,71 (3,96; 7,32)	89,06 (4,2; 6,73)
		α_1	85,29 (3,82; 10,88)	88,18 (4,02; 7,79)	88,86 (4,05; 7,08)	89,40 (4,23; 6,36)
		ϕ	91,02 (4,18; 4,79)	90,33 (4,81; 4,85)	89,81 (4,91; 5,27)	90,00 (4,57; 5,42)

Tabela 2.7 – Taxas de cobertura (%) empíricas e de não cobertura (para esquerda; para direita) dos intervalos de confiança, a níveis de confiança de 90% e 95%, dos parâmetros da distribuição GUI_{m01} com $\gamma = 0,5$ e $\phi = 30$.

α	$1 - \delta$	Parâmetro	n			
			40	80	120	200
$\alpha_0 = 0,24; \alpha_1 = 0,26$ ($\delta_0 = 12\%; \delta_1 = 13\%$)	95%	γ	94,66 (2,46; 2,87)	94,30 (3,0; 2,69)	94,74 (2,69; 2,56)	94,87 (2,51; 2,61)
		α_0	91,83 (1,84; 6,32)	92,56 (1,86; 5,57)	93,82 (1,72; 4,45)	94,22 (1,94; 3,83)
		α_1	90,98 (1,82; 7,19)	93,26 (2,14; 4,59)	94,06 (2,03; 3,9)	94,04 (2,23; 3,72)
		ϕ	95,83 (1,4; 2,76)	95,46 (1,77; 2,76)	95,23 (2,02; 2,74)	95,07 (2,21; 2,71)
		γ	89,45 (5,19; 5,35)	89,02 (5,74; 5,23)	89,65 (5,14; 5,2)	89,88 (5,06; 5,05)
		α_0	85,03 (3,8; 11,16)	88,00 (3,7; 8,29)	88,59 (3,91; 7,49)	89,18 (4,02; 6,79)
		α_1	87,34 (4,11; 8,54)	88,55 (4,1; 7,34)	88,93 (4,25; 6,81)	89,03 (4,59; 6,37)
		ϕ	91,33 (4,17; 4,49)	90,22 (4,84; 4,93)	90,57 (4,41; 5,01)	90,28 (4,84; 4,87)

As Tabelas 2.8 e 2.9 apresentam os resultados das simulações para os testes de hipóteses em alguns dos cenários considerados. As hipóteses testadas foram $\mathcal{H}_0 : \gamma_0 = 0,8$ versus $\mathcal{H}_1 : \gamma_0 \neq 0,8$ (Tabela 2.8) e $\mathcal{H}_0 : \gamma_0 = 0,5$ versus $\mathcal{H}_1 : \gamma_0 \neq 0,5$ (Tabela 2.9), respectivamente. Para a distribuição GUI_{m01} também testamos a hipótese $\mathcal{H}_0 : \alpha_{10} = 0,1$ versus $\mathcal{H}_1 : \alpha_{10} \neq 0,1$ (Tabela 2.9). É possível observar que as taxas empíricas de rejeição nula convergem para as taxas nominais correspondentes conforme o tamanho da amostra aumenta. Em geral, o teste de Wald apresentou o pior desempenho, ou seja, é o teste com o tamanho mais distorcido quando $n = 80, 120, 200$ (Tabela 2.8).

Por exemplo, para $\gamma_0 = 0,8$, $\phi = 30$ e nível de significância de 5% (Tabela 2.8), a taxa de rejeição nula para o teste de Wald foi de 5,36%, enquanto para o teste da razão de verossimilhança, escore e gradiente foram, respectivamente, 5,04%, 4,97% e 4,87%. Também observamos esse comportamento no cenário de dupla inflação (Tabela 2.9), de forma mais acentuada quando testamos o parâmetro γ_0 . Os testes tornam-se mais precisos quando $n = 200$. Considere, por exemplo, $\gamma_0 = 0,8$ e nível de significância de 10% (Tabela 2.8), os tamanhos dos testes para razão de verossimilhança, escore e gradiente foram, respectivamente, 9,95%, 10,06% e 9,96%.

Tabela 2.8 – Taxas de rejeição nula para os testes razão de verossimilhança (RV), Wald (W), escore (S_R) e gradiente (S_G). Distribuição GUI_{m1} . $\mathcal{H}_0: \gamma = \gamma_0 \times \mathcal{H}_1: \gamma \neq \gamma_0$. $\gamma = 0,8$, $\phi = 30$ e $\alpha_1 = 0,1$.

γ_0	δ	Teste	n			
			40	80	120	200
0,8	1%	RV	0,58	1,01	1,03	1,07
		W	1,01	1,98	1,59	1,51
		S_R	1,02	1,00	0,93	1,08
		S_G	0,98	0,97	0,92	1,01
	5%	RV	3,75	5,40	5,13	5,04
		W	5,13	6,33	5,88	5,36
		S_R	3,63	4,89	4,97	4,97
		S_G	3,55	4,72	4,80	4,87
	10%	RV	8,12	10,32	10,58	9,95
		W	9,95	11,42	10,82	10,31
		S_R	7,46	9,84	10,46	10,06
		S_G	7,29	9,61	10,31	9,96

Tabela 2.9 – Taxas de rejeição nula para os testes razão da verossimilhança (RV), Wald (W), escore (S_R) e gradiente (S_G). Distribuição GUI_{m01} . $\mathcal{H}_0: \gamma = \gamma_0 \times \mathcal{H}_1: \gamma \neq \gamma_0$. $\gamma = 0,5$, $\phi = 12$ ($\delta_0 = 13\%$; $\delta_1 = 12\%$) e $\mathcal{H}_0: \alpha_1 = \alpha_{10} \times \mathcal{H}_1: \alpha_1 \neq \alpha_{10}$. $\gamma = 0,5$, $\phi = 12$ ($\delta_0 = 13\%$; $\delta_1 = 12\%$).

γ_0	δ	Teste	n			
			40	80	120	200
0,5	1%	RV	0,88	1,05	1,06	1,08
		W	1,69	1,64	1,72	1,79
		S_R	0,26	0,52	0,51	0,55
		S_G	0,61	0,90	0,96	0,97
	5%	RV	5,22	5,61	5,13	5,07
		W	7,10	7,27	6,92	6,88
		S_R	2,97	3,46	3,54	3,45
		S_G	4,40	5,17	4,88	4,96
	10%	RV	10,63	10,72	10,16	10,22
		W	13,09	13,06	12,59	12,75
		S_R	7,38	8,00	7,77	7,72
		S_G	9,77	10,39	9,95	10,06
α_{10}	1%	RV	0,50	1,22	1,09	1,21
		W	3,35	2,02	1,59	1,40
		S_R	1,29	1,41	1,64	1,58
		S_G	0,69	0,86	0,97	1,10
	5%	RV	5,35	5,06	5,28	5,11
		W	6,77	5,84	5,06	4,73
		S_R	5,57	6,31	6,31	6,46
		S_G	4,14	4,96	4,97	5,17
	10%	RV	9,24	10,21	10,06	9,96
		W	12,74	10,13	9,38	8,96
		S_R	9,94	12,11	11,71	12,12
		S_G	8,26	9,83	9,62	9,76

2.4 APLICAÇÕES

Nesta seção, ilustraremos a aplicação da distribuição gama unitária inflacionada em conjunto de dados reais. Na primeira aplicação, utilizaremos a distribuição gama unitária inflacionada em um, no ajuste de dados de acesso à rede de esgotamento sanitário no estado de São Paulo. Na sequência, trazemos uma aplicação da distribuição gama unitária inflacionada em zero e um, utilizando o índice de esgoto tratado em relação à água consumida no estado de São Paulo. Mais detalhes sobre as aplicações são dados a seguir.

2.4.1 Dados de acesso à rede de esgotamento sanitário para o Estado de São Paulo

A variável de interesse é a proporção da população urbana residente em domicílios ligados à rede de esgotamento sanitário para cada município do estado de São Paulo (y). A fonte de dados utilizada foi o Atlas do Desenvolvimento Humano no Brasil (<www.atlasbrasil.org.br>). Os dados referem-se ao ano de 2017. Essa variável assume valores no intervalo $(0,1]$, apresentando 255 observações iguais a um (41,53%) de um total de 614 municípios. A Tabela 2.10 apresenta algumas estatísticas descritivas para a variável observada. Nota-se que a proporção média da população residente em domicílios com rede de esgotamento sanitário é de 0,9160 e que o mínimo observado foi de 0,1504. Além disso, pode-se destacar que 50% dos municípios observados possuem 99,12% da população urbana residente em domicílios com acesso à rede de esgotamento sanitário.

Tabela 2.10 – Estatísticas descritivas da variável proporção da população urbana residente em domicílios ligados à rede de esgotamento sanitário (y).

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão	CV
0,1504	0,9102	0,9912	0,9160	1,0000	1,0000	0,1544	0,1686

Ajustamos a distribuição GUI_{m1} e as estimativas de máxima verossimilhança dos parâmetros que indexam essa distribuição são $\hat{\alpha}_1 = 0,4549$, $\hat{\gamma} = 0,9130$ e $\hat{\phi} = 0,5361$. Usando a estimativa de máxima verossimilhança dos parâmetros, construímos intervalos de confiança assintóticos com 95% de confiança. Assim, os respectivos intervalos para γ , α_1 e ϕ foram $(0,9126; 0,9135)$, $(0,4533; 0,4565)$ e $(0,5335; 0,5388)$. Na Figura 2.7 são apresentados o histograma, além dos gráficos da função de distribuição empírica e a acumulada ajustada considerando o modelo GUI_{m1} . Através do gráfico é possível perceber que a distribuição GUI_{m1} se ajusta bem aos dados.

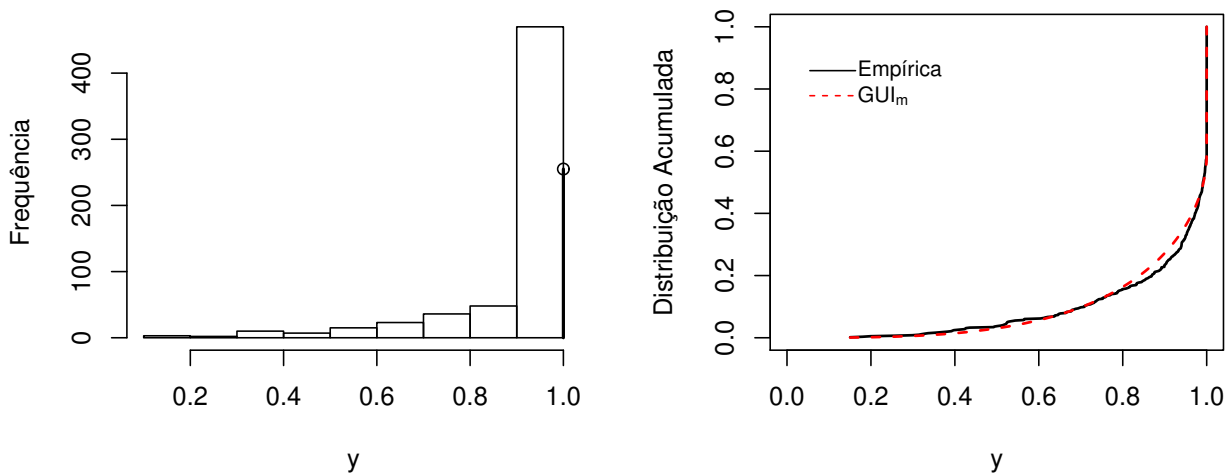


Figura 2.7 – Histograma e função de distribuição acumulada para a variável percentual da população em domicílios ligados à rede de esgotamento sanitário.

Por fim, para avaliar a performance dos testes de hipóteses, levamos em consideração algumas informações da PNAD contínua - características gerais dos domicílios e dos moradores - em que apresenta um levantamento acerca da situação dos domicílios ligados à rede de esgoto para o Brasil e regiões. Em alguns dos resultados, no que diz respeito a proporção de domicílios com acesso à rede geral de esgotos, a pesquisa destacou que a região Sudeste alcançou uma estimativa de 88,9% em 2019 (Pesquisa Nacional por Amostra de Domicílios Contínua, 2020). Neste sentido, as hipóteses nulas versus as hipóteses alternativas bilaterais correspondentes para testar a média foram definidas como: $\mathcal{H}_0 : \gamma = 0,89$ versus $\mathcal{H}_1 : \gamma \neq 0,89$ e $\mathcal{H}_0 : \gamma = 0,90$ versus $\mathcal{H}_1 : \gamma \neq 0,90$. A Tabela 2.11 apresenta os p -valores para os testes realizados. Nota-se que os testes rejeitaram as hipóteses nulas, ao nível de significância de 5%. Isto é, há evidência estatística de que o percentual da população residente em domicílios ligados à rede de esgotamento sanitário difere de 89% e 90%.

Tabela 2.11 – Resultados dos testes de hipóteses: razão de verossimilhança (RV), Wald (W), escore (S_R) e gradiente (S_G) para testar diferentes valores de γ .

	Teste	Estatística	p -valor
$\mathcal{H}_0 : \gamma = 0,89$	RV	12,3170	0,0004
	W	15,5185	0,0001
	S_R	10,7643	0,0010
	S_G	11,0015	0,0009
$\mathcal{H}_0 : \gamma = 0,90$	RV	4,3347	0,0373
	W	4,9619	0,0259
	S_R	4,0071	0,0453
	S_G	4,0551	0,0440

Com o objetivo de avaliar o desempenho da distribuição proposta com outras distribuições inflacionadas já existentes na literatura, realizamos uma comparação a partir das medidas de bon-

dade de ajuste: critério de informação de Akaike (AIC), critério de informação de Akaike corrigido (AICc), critério de informação bayesiano (BIC) e critério de informação de Hannan-Quinn (HQIC), no que se refere à adequabilidade do modelo. Neste sentido, quanto menor o valor das estatísticas, melhor o ajuste da distribuição aos dados. Mais detalhes sobre essas medidas podem ser consultados em Akaike (1998), Schuwarz (1978), Sugiura (1978), Hannan e Quinn (1979). Os resultados são sumarizados na Tabela 2.12. Para todas as medidas analisadas há uma superioridade da distribuição proposta, configurando-a como uma alternativa competitiva com as demais distribuições consideradas. Além disso, vale ressaltar que na distribuição GUI_m temos a média da distribuição inflacionada, ao contrário das outras distribuições, em que a estimativa da média e/ou mediana refere-se apenas a parte contínua do modelo. O conjunto de dados utilizado na primeira aplicação apresenta 41,53% de valores iguais a 1 e estão sendo considerados no momento que estimamos a média a partir da distribuição proposta, ao passo que com as distribuições beta inflacionada (OSPINA; FERRARI, 2010) e Kumaraswamy inflacionada isso não é possível. Por exemplo, a média estimada pelo modelo proposto foi igual a $\hat{\gamma} = 0,9130$, enquanto a média estimada a partir da distribuição beta inflacionada foi $\hat{\mu} = 0,8511$. Assim, no primeiro caso, os municípios que apresentam 100% de acesso à rede de esgotamento sanitário foram considerados quando estimamos a média, o que leva a uma estimativa mais apropriada para descrever o fenômeno analisado.

Tabela 2.12 – Critérios de adequabilidade de modelos: AIC, AICc, BIC e HQIC - dados de acesso à rede de esgotamento sanitário para o estado de São Paulo.

Distribuição	AIC	AICc	BIC	HQIC
Gama unitária inflacionada	103,2528	103,2920	116,5176	108,4108
Beta inflacionada	103,4795	103,5188	116,7444	108,6375
Kumaraswamy inflacionada	105,6174	105,6567	118,8823	110,7754

2.4.2 Proporção de esgoto tratado em relação à água consumida no Estado de São Paulo

Nesta aplicação, utilizamos a proporção de esgoto tratado em relação à água consumida para os municípios do estado de São Paulo. Sob essa abordagem, a variável de interesse pode assumir os valores zero e um. A amostra foi composta por 635 municípios e a quantidade de municípios com índice de esgoto tratado igual a 0 foi de 38 (6,0%) e igual a 1 sendo 122 (19,21%). O conjunto de dados foi obtido no Sistema Nacional de Informação sobre Saneamento (<www.gov.br/mdr/pt-br/assuntos/saneamento/pmss/snis>) e as informações referem-se ao ano de 2020. A partir da Tabela 2.13 nota-se que o índice médio de esgoto tratado para os municípios foi de 0,7598, com desvio padrão de 0,2989. Ainda, verifica-se a existência de alguns municípios em que o índice de esgoto tratado é igual a zero.

Segundo as diretrizes nacionais para o saneamento básico, os contratos de prestação dos serviços públicos de saneamento básico devem definir metas de universalização que assegurem o atendimento de 90% da população com coleta e tratamento de esgotos. No Brasil, entre os anos de 2016 a 2020, observa-se que a proporção de esgoto tratado em relação à água consumida vem aumentando, passando de 44,9%(2016) para 50,8%(2020). Com relação as macrorregiões do país, apenas a região Sudeste e Centro-Oeste tratam mais da metade do esgoto gerado, sendo de 58,6% e 58,5%, respecti-

vamente (Secretaria Nacional de Saneamento - SNS, 2022). O Estado de São Paulo apresenta 8 dos 20 municípios mais bem colocados no Ranking de saneamento básico, segundo o relatório do Trata Brasil 2022. Os municípios de Piracicaba e Jundiaí estão entre os que tiveram 100% de volume de esgoto tratado em relação à água consumida (Instituto Trata Brasil, 2022).

Tabela 2.13 – Estatísticas descritivas da variável proporção de esgoto tratado referido à água consumida para os municípios do Estado de São Paulo.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio padrão	CV
0,0000	0,6630	0,8820	0,7598	0,9877	1,0000	0,2989	0,3934

Em seguida, ajustamos a distribuição GUI_{m01} cujas estimativas de máxima verossimilhança para os parâmetros da distribuição são $\hat{\alpha}_0 = 0,2367$, $\hat{\alpha}_1 = 0,2571$, $\hat{\gamma} = 0,7471$ e $\hat{\phi} = 0,5986$. Aqui, diferentemente da aplicação anterior, foi estimado mais um parâmetro (α_0) da distribuição GUI_m em decorrência dos valores zeros observados na amostra. Utilizando as estimativas de máxima verossimilhança, os intervalos de confiança para os parâmetros α_0 , α_1 , γ e ϕ foram, respectivamente, (0,2342; 0,2391), (0,2556; 0,2587), (0,7462; 0,7481) e (0,5961; 0,6012). Com relação aos testes de hipóteses, foram consideradas as seguintes hipóteses nulas para testar a média: $\mathcal{H}_0 : \gamma = 0,75$ versus $\mathcal{H}_1 : \gamma \neq 0,75$ e $\mathcal{H}_0 : \gamma = 0,90$ versus $\mathcal{H}_1 : \gamma \neq 0,90$. Conforme os resultados, não rejeitamos a hipótese nula de a proporção média de esgoto tratado em relação à água consumida no estado São Paulo é igual a 75%, conforme Tabela 2.14. De outro modo, quando testamos $\mathcal{H}_0 : \gamma = 0,90$, os resultados dos testes assinalam para a rejeição da hipótese nula, ao nível nominal de 5%. Na Figura 2.8 são apresentados o histograma, bem como os gráficos da função de distribuição empírica e a acumulada ajustada considerando o modelo GUI_{m01} . É possível observar que a distribuição GUI_{m01} ajusta-se de forma adequada aos dados observados.

Posteriormente, considerando alguns critérios de adequabilidade de modelos, comparamos a distribuição GUI_{m01} com as distribuições beta inflacionada e Kumaraswamy inflacionada (Tabela 2.15). Conforme os resultados, nota-se que a distribuição GUI_{m01} destaca-se, de modo geral, em relação as demais distribuições inflacionadas. Observamos que as diferenças são mais expressivas quando temos a distribuição proposta e a distribuição Kumaraswamy inflacionada. Adicionalmente, torna-se importante destacar que a proporção média de esgoto tratado referido à água consumida estimada a partir da distribuição proposta leva em conta os municípios que apresentaram proporção igual a zero e a um (isto é, 100% de esgoto tratado), ao contrário das distribuições beta inflacionada e Kumaraswamy inflacionada, em que essas observações não são incluídas na estimação. Essa perda de informação pode causar impacto nos resultados obtidos, uma vez que o modelo não incorpora a variável com todas as suas características e, principalmente, quando o número de observações iguais a zero e/ou um na amostra for elevado.

Tabela 2.14 – Resultados dos testes de hipóteses: razão de verossimilhança (RV), Wald (W), escore (S_R) e gradiente (S_G) para testar diferentes valores de γ .

	Teste	Estatística	p -valor
$\mathcal{H}_0 : \gamma = 0,75$	RV	0,0605	0,8057
	W	0,0620	0,8034
	S_R	0,0588	0,8084
	S_G	0,0607	0,8053
$\mathcal{H}_0 : \gamma = 0,90$	RV	321,8309	< 0,0000
	W	178,2608	< 0,0000
	S_R	465,5057	< 0,0000
	S_G	466,0203	< 0,0000

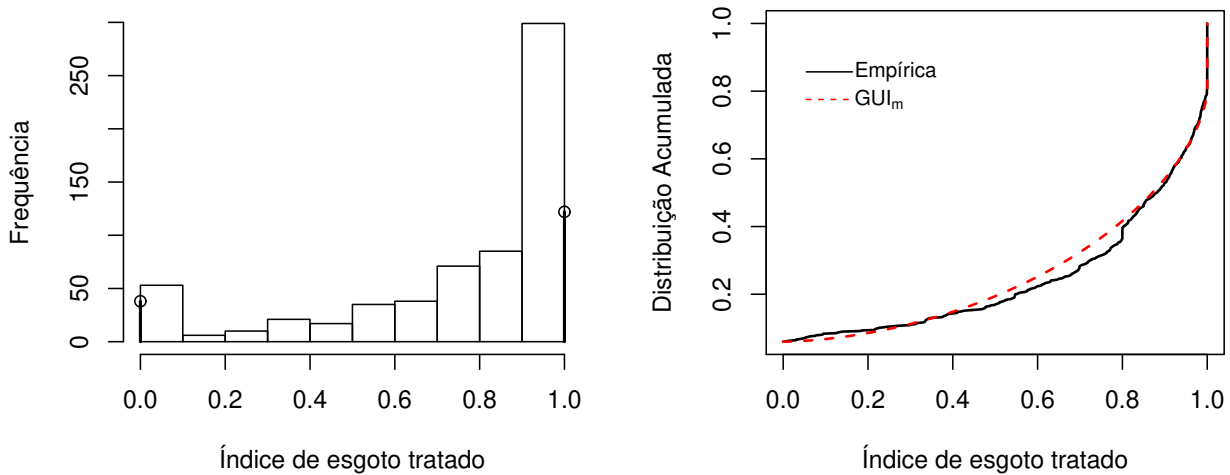


Figura 2.8 – Histograma e função de distribuição acumulada do índice de esgoto tratado.

Tabela 2.15 – Critérios de adequabilidade de modelos: AIC, AICc, BIC e HQIC - Índice de esgoto tratado referido à água consumida no estado de São Paulo.

Distribuição	AIC	AICc	BIC	HQIC
Gama unitária inflacionada	477,7881	477,8516	495,6026	484,7053
Beta inflacionada	479,6050	479,6685	497,4195	486,5222
Kumaraswamy inflacionada	482,5946	482,6581	500,4091	489,5117

2.5 CONCLUSÃO

Neste artigo propomos a distribuição gama unitária inflacionada para modelar dados restritos aos intervalos $[0,1]$, $(0,1]$ ou $[0,1)$. A estimação dos parâmetros foi derivada a partir do método de máxima verossimilhança. Realizamos um estudo de simulação para avaliar o desempenho dos estimadores pontuais e intervalares, além da performance de testes de hipóteses assintóticos para os parâmetros da distribuição. Apresentamos duas aplicações a dados reais para demonstrar o uso da distribuição proposta. Os resultados obtidos assinalam que a distribuição GUI_m pode ser considerada como uma alternativa competitiva para a modelagem de dados nos intervalos $(0,1]$, $[0,1)$ ou $[0,1]$. Além

disso, comparamos a distribuição proposta com as distribuições beta e Kumaraswamy inflacionadas através de alguns critérios de informação. Observamos, em geral, que a distribuição gama unitária inflacionada apresentou performance superior aos demais modelos inflacionados considerados. Em alguns momentos, destacando-se de forma mais expressiva em relação a distribuição Kumaraswamy inflacionada, como observado na segunda aplicação.

Adicionalmente, vale salientar que há duas vantagens significativas quanto ao uso da parametrização proposta. A primeira sendo a possibilidade de modelar a média da variável resposta, tornando-se, por sua vez, mais atrativa para o uso em modelos de regressão e gráficos de controle. E a segunda vantagem está relacionada a formulação do modelo. Isto é, a partir de uma única formulação é possível obter a distribuição gama unitária inflacionada em zero, um ou zero e um. Dessa forma, recomendamos o uso da distribuição proposta para modelar o comportamento de variáveis que seguem distribuição gama unitária e apresentam zeros e/ou uns com probabilidade positiva.

3 GRÁFICO DE CONTROLE DO TIPO SHEWHART CONSIDERANDO A DISTRIBUIÇÃO GAMA UNITÁRIA INFLACIONADA EM ZERO OU UM

RESUMO: O presente artigo propõe um gráfico de controle para monitorar características de qualidade que assumem valores nos intervalos $(0,1]$ ou $[0,1)$, como taxas ou proporções, que podem conter o valor zero ou um com probabilidade positiva. O gráfico proposto tem os limites de controle definidos a partir da distribuição gama unitária inflacionada proposta no capítulo anterior. Um extenso estudo de simulação de Monte Carlo foi realizado para avaliar a performance dos gráficos de controle em termos de comprimento da sequência. Realizamos uma comparação entre o gráfico de controle proposto com o gráfico de controle beta inflacionado, considerando duas abordagens. Na primeira, consideramos observações individuais e na segunda, subgrupos amostrais, s , de tamanhos 8, 15, 30 e 50. Em gráficos de controle, quando o tamanho do subgrupo é maior que 1, é comum o interesse em monitorar a média da característica de qualidade. Os resultados numéricos evidenciam que o gráfico proposto apresentou bom desempenho nas duas abordagens apresentadas, isto é, considerando observações individuais e não individuais. Adicionalmente, duas aplicações a conjunto de dados reais ilustram a aplicabilidade do gráfico de controle proposto.

Palavras-chave: Gráfico de controle, distribuição gama unitária inflacionada, taxas e proporções, controle estatístico de processos.

ABSTRACT: This study proposes a control chart to monitor quality characteristics that take values in the intervals $(0,1]$ or $[0,1)$, such as fractions or proportions, which may contain the value zero or one with positive probability. The proposed graph has the control limits defined from the inflated unit gamma distribution proposed in the previous chapter. An extensive Monte Carlo simulation study was conducted to evaluate the performance of the control charts in terms of sequence length. We perform a comparison between the proposed control chart and the inflated beta control chart, considering two approaches. Firstly, we consider individual observations and secondly, sample subgroups of size $s = 8, 15, 30$ e 50 . In control charts, when the subgroup size is larger than 1, it is common to be interested in monitoring the mean of the quality characteristic. The numerical results show that the proposed graph performed well in the two approaches presented, that is, considering individual and non-individual observations. Additionally, two applications to real data sets illustrate the applicability of the proposed control chart.

Keywords: Control chart, inflated unit gamma distribution, rates and proportions, statistical process control.

3.1 INTRODUÇÃO

O controle estatístico de qualidade é um campo da estatística que envolve algumas áreas, entre elas, o controle estatístico de processo, CEP, que tem se tornado uma ferramenta essencial no contexto de controle de qualidade, ocupando um papel importante no controle e na melhoria da qualidade e produtividade de processos industriais (BACKER; BROBST, 1996; DUARTE; SARAIVA, 2008). Os gráficos de controle, uma das ferramentas do CEP, foram originalmente desenvolvidos para monitorar processos industriais (MONTGOMERY, 2009). No entanto, sua aplicabilidade estende-se além desses processos. Entre algumas aplicações, destacam-se o monitoramento da qualidade da água (PAROISSIN et al., 2016; LIN et al., 2020), abastecimento de água (GOVE et al., 2013), temperatura (LUND; SEYMOUR, 1999), umidade relativa (LIMA-FILHO et al., 2023), níveis perigosos de contaminantes (LEIVA et al., 2015) e o monitoramento de sistemas de produção animal (VRIES; RENEAU, 2010).

É comum termos o interesse em monitorar variáveis contínuas que assumem valores no intervalo $(0,1)$, como taxas e proporções. Algumas distribuições que acomodam variáveis dessa natureza têm sido utilizadas para o monitoramento de processos em diversas aplicações. Sant'anna e Caten (2012) propuseram um gráfico de controle beta, baseado na distribuição de probabilidade beta (GUPTA; NADARAJAH, 2004), para monitorar processos duplamente limitados. Ho et al. (2019) compararam o desempenho do gráfico de controle beta, com outros gráficos de controle derivados a partir das distribuições simplex e gama unitária. Adicionalmente, Lima-Filho e Bayer (2021) propuseram um gráfico de controle para o monitoramento de dados ambientais restritos ao intervalo $(0,1)$, com base na distribuição Kumaraswamy (KUMARASWAMY, 1980). E ainda, Lima-Filho et al. (2020a) propuseram gráficos de controle para monitorar características da qualidade assimétricas duplamente limitadas. Nesta última abordagem, os autores consideram as distribuições Kumaraswamy e Weibull unitária para a construção dos gráficos e comparam diferentes estimadores para o parâmetro da mediana. Todavia, em alguns casos, dados como frações e proporções podem conter os valores zero e/ou um e, dessa forma, os gráficos de controle descritos anteriormente não são adequados para variáveis com essa característica.

Na modelagem de variáveis duplamente limitadas, que assumem valores zero e/ou um, Lima-Filho et al. (2019) propuseram o gráfico de controle beta inflacionado, que é construído com base na distribuição beta inflacionada proposta por Bayes e Valdivieso (2016). Além disso, neste contexto, também temos o gráfico de controle Kumaraswamy inflacionado, proposto por Lima-Filho et al. (2023). Os autores propuseram o gráfico de controle baseado na distribuição Kumaraswamy inflacionada para monitorar o parâmetro da mediana.

O desenvolvimento de gráficos para o monitoramento de processos a partir de distribuições inflacionadas no intervalo unitário ainda é escasso. Neste sentido, métodos baseados em novas distribuições inflacionadas podem ser relevantes nas situações em que os modelos já propostos não se adequem, podendo conferir maior flexibilidade na modelagem dessas variáveis. A distribuição gama

unitária proposta por Grassia (1977), é adequada para modelar variáveis do tipo taxas e proporções. Uma característica interessante dessa distribuição é sua flexibilidade em acomodar diferentes formas. Nos últimos anos ela tem ganhado especial atenção de alguns pesquisadores em trabalhos que tiveram como destaque: aspectos de melhoramento de testes de hipóteses (GUEDES et al., 2019), métodos de estimação dos parâmetros da distribuição gama unitária (MAZUCHELI et al., 2018b), modelos de regressão (MOUSA et al., 2016) e medidas de diagnóstico em modelos de regressão gama unitária (ROCHA et al., 2021).

A distribuição gama unitária proposta no capítulo anterior é baseada em uma parametrização em que é possível estimar a média da distribuição inflacionada e não apenas a média referente à componente contínua do modelo, como visto nas distribuições inflacionadas usuais. Essa abordagem se torna mais interessante na construção de modelos de regressão e gráficos de controle, uma vez que a finalidade é estimar a média do modelo ou do processo, respectivamente. Assim, com base na distribuição gama unitária inflacionada apresentada, propomos o gráfico de controle gama unitária inflacionado.

O artigo se encontra estruturado da seguinte forma: na Seção 3.2 apresentamos alguns aspectos relacionados ao gráfico de controle gama unitária inflacionado. Na Seção 3.3 trazemos a forma de avaliação de desempenho do gráfico de controle e os cenários considerados no estudo de simulação. As Seções 3.4 e 3.5 trazem os resultados numéricos do estudo de simulação de Monte Carlo, a partir de duas abordagens, nomeadas de medidas individuais e não individuais. Na Seção 3.6 apresentamos duas aplicações a banco de dados reais para ilustrar a aplicabilidade do gráfico proposto. E por fim, na Seção 3.7, descrevemos as principais conclusões do artigo.

3.2 GRÁFICO DE CONTROLE GAMA UNITÁRIA INFLACIONADO

A distribuição gama unitária, GU, (GRASSIA, 1977) é apropriada para modelar variáveis que assumem valores no intervalo unitário padrão $(0,1)$. Por outro lado, quando consideramos variáveis que podem assumir valores em algum dos intervalos $[0,1)$, $(0,1]$ ou $[0,1]$, a distribuição GU torna-se inadequada para a modelagem. A distribuição gama unitária inflacionada estende a classe de distribuições gama unitária, possibilitando a modelagem de variáveis com inflação de zeros e/ou uns. A distribuição é composta por uma mistura entre uma distribuição contínua (gama unitária) e uma distribuição discreta, a saber: Bernoulli.

Lima-Filho et al. (2019) propuseram o gráfico de controle beta inflacionado, baseado em uma reparametrização apresentada por Bayes e Valdivieso (2016), em que é possível modelar a média da distribuição beta inflacionada e não apenas a média da distribuição beta. Os autores evidenciam a superioridade do gráfico de controle beta inflacionado em relação ao gráfico de controle beta (SANT'ANNA; CATEN, 2012). Estudos voltados a esse contexto ainda são escassos, fazendo com que o monitoramento de variáveis definidas nos intervalos $(0,1]$ ou $[0,1)$ seja, de certa forma, limitado. Assim, torna-se pertinente o desenvolvimento de gráficos direcionados ao monitoramento de

variáveis dessa natureza, tendo em vista a restrição de modelos propostos para tal fim. A partir da reparametrização apresentada, é possível modelar a média da distribuição gama unitária inflacionada, ao invés da média da distribuição gama unitária, que não considera os valores iguais a zero ou um. Essa abordagem é mais atrativa para o monitoramento de processos, diferente de outras distribuições inflacionadas propostas na literatura (OSPINA; FERRARI, 2010; LIU et al., 2020b).

3.2.1 Distribuição gama unitária inflacionada

Nesta seção apresentamos a distribuição gama unitária inflacionada (GUI_m), bem como o gráfico de controle para medidas individuais ($s = 1$) e medidas não individuais ($s > 1$) baseados nesta distribuição. Suponha que uma variável aleatória Y segue distribuição gama unitária inflacionada. A função densidade de probabilidade com parâmetros γ , α_0 , α_1 e ϕ , isto é, $Y \sim GUI_m(\gamma, \alpha_0, \alpha_1, \phi)$, é dada por:

$$f(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} \alpha_0(1 - \gamma), & \text{se } y = 0 \\ (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)gu(y; \check{\mu}, \phi), & \text{se } y \in (0, 1) \\ \alpha_1\gamma, & \text{se } y = 1, \end{cases} \quad (10)$$

em que $gu(y; \check{\mu}; \phi)$ é a função densidade da distribuição gama unitária, com $\check{\mu} = \frac{\gamma(1-\alpha_1)}{1-\alpha_0(1-\gamma)-\alpha_1\gamma}$, γ , $\alpha_0, \alpha_1 \in (0, 1)$ e $\phi > 0$. Note que $P(Y = 0) = \alpha_0(1 - \gamma)$ e $P(Y = 1) = \alpha_1\gamma$. Ainda, com base nessa parametrização, temos como casos particulares a distribuição inflacionada em zero, isto é, se $\alpha_0 > 0$ e $\alpha_1 = 0$; a distribuição gama unitária inflacionada em um, se $\alpha_1 > 0$ e $\alpha_0 = 0$, e por último, se $\alpha_0 = 0$ e $\alpha_1 = 0$, temos a distribuição gama unitária.

A média e a variância associadas a função densidade de probabilidade dada em (10) são, respectivamente:

$$E(Y) = \gamma,$$

e

$$Var(Y) = \alpha_1\gamma(1 - \alpha_1\gamma) + \gamma(1 - \alpha_1) \left[\left(2 - \left(\frac{\gamma(1 - \alpha_1)}{1 - \alpha_0(1 - \gamma) - \alpha_1\gamma} \right)^{1/\phi} \right)^{-\phi} - \gamma(1 + \alpha_1) \right].$$

A função de distribuição acumulada de uma variável aleatória $Y \sim GUI_m(\gamma, \alpha_0, \alpha_1, \phi)$ é dada por:

$$F(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} 0, & \text{se } y < 0 \\ \alpha_0(1 - \gamma), & \text{se } y = 0 \\ \alpha_0(1 - \gamma) + (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)GU(y; \check{\mu}, \phi), & \text{se } 0 < y < 1 \\ 1, & \text{se } y \geq 1. \end{cases} \quad (11)$$

Quando temos um processo sob controle, isto é, os valores da característica de qualidade estão todos dentro do limite do intervalo, não há motivos para se tomar medidas corretivas. Quando identificamos alguma mudança no processo, é necessário uma análise mais detalhada afim de determinar a causa para a mudança de comportamento. A seguir apresentamos a construção dos gráficos de controle para monitorar medidas individuais e não individuais a partir da distribuição GUI_m .

3.2.2 Gráfico de controle para monitorar medidas individuais

Seja Y uma variável aleatória que segue a densidade dada em (10), com média γ , precisão ϕ e os parâmetros α_0 e α_1 . Considere uma região de controle $(1 - \alpha)$, em que α representa a probabilidade de Y não estar entre os limites inferior e superior do gráfico (Erro tipo I). Assim, podemos escrever

$$P_0(LIC|\alpha_{00}, \alpha_{10}, \gamma_0, \phi_0 \leq Y \leq LSC|\alpha_{00}, \alpha_{10}, \gamma_0, \phi_0) = 1 - \alpha, \quad (12)$$

em que P_0 é a probabilidade calculada quando o processo está sob controle, $\alpha_{00}, \alpha_{10}, \gamma_0$ e ϕ_0 são os parâmetros do processo sob controle, LIC é o limite inferior de controle e LSC o limite superior de controle. Por simplicidade, os limites de controle podem ser expressos como $P_0(Y \leq LIC|\gamma_0, \alpha_{00}, \alpha_{10}, \phi_0) = P_0(Y \geq LSC|\gamma_0, \alpha_{00}, \alpha_{10}, \phi_0) = \frac{\alpha}{2}$. Neste sentido, os limites de controle baseados na distribuição gama unitária inflacionada podem ser definidos como

$$LIC = F^{-1}(\alpha/2; \alpha_{00}, \alpha_{10}, \gamma_0, \phi_0),$$

$$LC = \gamma_0,$$

$$LSC = F^{-1}(1 - \alpha/2; \alpha_{00}, \alpha_{10}, \gamma_0, \phi_0),$$

em que $F^{-1}(\cdot)$ é a função quantil da distribuição gama unitária inflacionada.

O poder de detecção, calculado a partir da probabilidade do erro tipo II (β), considerando uma mudança δ na média do processo ($\gamma_1 = \gamma_0 \times \delta$), é calculado como

$$1 - \beta = P_1(Y < LIC|\alpha_{00}, \alpha_{10}, \gamma_1, \phi_0) + P_1(Y > LSC|\alpha_{00}, \alpha_{10}, \gamma_1, \phi_0),$$

em que P_1 é a probabilidade calculada considerando que a média do processo está fora de controle. Na prática, os parâmetros sob controle $\alpha_{00}, \alpha_{10}, \gamma_0$ e ϕ_0 são desconhecidos e, dessa forma, precisam ser estimados. Os estimadores de máxima verossimilhança (EMV) são amplamente utilizados (PAWITAN, 2001; CASELLA; BERGER, 2002) devido suas boas propriedades, entre elas a consistência e a normalidade assintótica.

3.2.3 Gráfico de controle \bar{X}

Em situações práticas, é possível obter subamostras para o monitoramento da característica da qualidade. Cada amostra de $s > 1$ observações é um subgrupo, em que podemos ter n subgrupos. Esta seção é dedicada a apresentação do gráfico de controle para monitorar a média da variável de interesse,

a partir do estimador de máxima verossimilhança. Devido a dificuldade em obter expressões de forma fechada para a distribuição do estimador, é necessário recorrer a métodos de reamostragem. Desse modo, sugerimos o procedimento conforme Lima-Filho et al. (2020a), que utilizam a distribuição empírica sob controle dos estimadores com base em métodos de reamostragem. Considere o estimador $\hat{\gamma}$ e um α fixo, os limites de controle podem ser obtidos por:

$$P(\hat{\gamma} \leq LIC | \alpha_{00}, \alpha_{10}, \gamma_0, \phi_0) = P(\hat{\gamma} \geq LSC | \alpha_{00}, \alpha_{10}, \gamma_0, \phi_0) = \frac{\alpha}{2}.$$

Os limites de controle, LIC e LSC, são os quantis $\alpha/2$ e $1 - \alpha/2$ da distribuição empírica do estimador $\hat{\gamma}$. A estimação do vetor de parâmetros $\theta = (\alpha_0, \alpha_1, \gamma, \phi)^\top$ é realizada através do método de máxima verossimilhança. Considere uma amostra aleatória $\mathbf{y} = y_1, \dots, y_s$, de tamanho s . A função de log verossimilhança, $\ell(\theta)$, para o vetor de parâmetros $\theta = (\alpha_0, \alpha_1, \gamma, \phi)$, pode ser expressa como

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^s \mathbb{1}_{\{0\}}(y_i) \log[\alpha_0(1 - \gamma)] + \sum_{i=1}^s \mathbb{1}_{\{1\}}(y_i) \log[\alpha_1\gamma] \\ &+ \sum_{i=1}^s (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \log c + \sum_{i=1}^s (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \log gu(y_i; \check{\mu}, \phi), \end{aligned}$$

com $c = 1 - \alpha_0(1 - \gamma) - \alpha_1\gamma$ e

$$\log gu(y_i; \check{\mu}, \phi) = \phi \log \left(\frac{\check{\mu}^{1/\phi}}{1 - \check{\mu}^{1/\phi}} \right) - \log \Gamma(\phi) + \log y_i \left(\frac{\check{\mu}^{1/\phi}}{1 - \check{\mu}^{1/\phi}} - 1 \right) + (\phi - 1) \log \left[\log \left(\frac{1}{y_i} \right) \right],$$

em que $\log gu(y_i; \check{\mu}, \phi)$ é a log-densidade da distribuição gama unitária. Como as derivadas de $\ell(\theta)$ são funções não lineares, os estimadores de máxima verossimilhança são obtidos através de métodos de otimização não linear (PRESS, 2007; JORGE; STEPHEN, 2006). As expressões para as primeiras derivadas com relação a cada um dos parâmetros são apresentadas no Apêndice A.

Adicionalmente, dado um processo sob controle com os parâmetros α_{00} , α_{10} , γ_0 e ϕ_0 , os procedimentos para implementar o gráfico de controle proposto são dados por:

1. Gere 10.000 subgrupos de tamanho s a partir da distribuição GUI_m ;
2. Para cada subgrupo gerado i , $i = 1, \dots, 10.000$, estime a média considerando o estimador $\hat{\gamma}$;
3. Com base nas 10.000 réplicas do estimador $\hat{\gamma}$, a distribuição empírica do estimador é determinada e os quantis empíricos $q_{\alpha/2, \hat{\gamma}}$ e $q_{1-\alpha/2, \hat{\gamma}}$, respectivamente, são obtidos.

No passo 1, se os parâmetros sob controle forem desconhecidos, eles serão substituídos pelas suas respectivas estimativas de máxima verossimilhança. Desse modo, os limites de controle do gráfico proposto na presença de subgrupos podem ser expressos por

$$LIC = q_{\alpha/2, \hat{\gamma}},$$

$$LC = \hat{\gamma}_*,$$

$$LSC = q_{1-\alpha/2, \hat{\gamma}},$$

em que $q_{\alpha/2, \hat{\gamma}}$ e $q_{1-\alpha/2, \hat{\gamma}}$ são os quantis empíricos da distribuição do estimador $\hat{\gamma}$ e $\hat{\gamma}_*$ é a média da distribuição empírica.

O poder de detecção, calculado a partir da probabilidade do erro tipo II (β), é determinado de forma similar ao caso em que temos medidas individuais.

3.3 AVALIAÇÃO DE DESEMPENHO DOS GRÁFICOS DE CONTROLE

O desempenho de um gráfico de controle é geralmente medido em termos do número de observações (comprimento da sequência) até que um sinal de alerta seja observado. O comprimento da sequência (*runing length*) é uma variável aleatória com distribuição geométrica de parâmetro p . Neste sentido, uma das medidas utilizadas é o comprimento médio da sequência (*average runing length* - ARL) (MONTGOMERY, 2009). O ARL fornece o número médio de observações até a detecção de uma condição fora de controle. A avaliação de desempenho é realizada mensurando o ARL em duas situações. A primeira, na presença de um processo sob controle, onde temos o ARL_0 . O ARL_0 é definido conforme a probabilidade do erro tipo I (α), com $ARL_0 = \frac{1}{\alpha}$, em que $\alpha = 1 - P_0(LIC|\alpha_{00}, \alpha_{10}, \gamma_0, \phi_0 \leq Y \leq LSC|\alpha_{00}, \alpha_{10}, \gamma_0, \phi_0)$. Um ARL_0 maior indica uma probabilidade menor de alarmes falsos (MONTGOMERY, 2009; CASTAGLIOLA; MARAVELAKIS, 2011). Por outro lado, quando temos um processo fora de controle, com média $\gamma_1 = \gamma_0 \times \delta$, em que γ_0 é a média considerando um processo sob controle e δ uma perturbação na média, o ARL_1 pode ser escrito como $\frac{1}{1-\beta}$, em que $\beta = P_1(LIC|\alpha_{00}, \alpha_{10}, \gamma_1, \phi_0 \leq Y \leq LSC|\alpha_{00}, \alpha_{10}, \gamma_1, \phi_0)$, sendo P_1 a probabilidade calculada considerando o processo fora de controle. Portanto, assumimos que o processo está fora de controle quando temos uma mudança de tamanho δ na média do processo, isto é, $\gamma_1 = \gamma_0 \times \delta$, enquanto para $\delta = 1$, o processo está sob controle. Vale destacar que o ARL_1 fornece o poder do gráfico em detectar mudanças no processo.

Para efeitos de avaliação de desempenho, duas medidas também importantes e frequentemente usadas nesse contexto foram consideradas, sendo elas a mediana da distribuição do comprimento (*Median Run Length* - MRL) e o desvio padrão da distribuição do comprimento (*Standard Deviation of Run Length* - SDRL) (KHALIQ et al., 2016; RIAZ, 2015; PAROISSIN et al., 2016; TEOH et al., 2017; AHMAD et al., 2019). As respectivas medidas podem ser definidas como $MRL = \ln(0,5)/\ln(1-p)$ e $SDRL = \sqrt{(1-p)/p^2}$ (HO et al., 2019), em que $p = \alpha$ quando o processo está sob controle e $p = 1 - \beta$ quando o processo está fora de controle (HO et al., 2019). Na avaliação de desempenho do gráfico de controle consideramos $\alpha = 0,0027$, abordagem semelhante a diversos estudos propostos na área (ASLAM et al., 2015; JOEKES et al., 2015; LOMBARD; MAXWELL, 2012; TEOH et al., 2017).

Na Tabela 3.1 são apresentados seis cenários considerados no estudo de simulação. Os ce-

nários 1, 2 e 3 consideram a distribuição gama unitária inflacionada em zero, enquanto os cenários 4, 5 e 6, consideram a distribuição gama unitária inflacionada em um. A reprodução de um cenário fora de controle foi realizada a partir de perturbações na média do processo ($\gamma_1 = \gamma_0 \times \delta$), em que $\delta = (1,05; 1,10; 1,15; 1,20; 1,25; 1,30; 1,35; 1,40; 1,45; 1,50)$ para os cenários inflacionados em zero e $\delta = (0,95; 0,90; 0,85; 0,80; 0,75; 0,70; 0,65; 0,60; 0,55; 0,50)$ para os cenários inflacionados em um. Salientando que $\delta = 1$ representa um processo sob controle e, neste caso, temos as medidas ARL_0 , MRL_0 e $SDRL_0$. Os valores esperados de ARL_0 , MRL_0 e $SDRL_0$ considerando $\alpha = 0,0027$ são, respectivamente, 370, 256 e 369. Todas as implementações computacionais e simulações foram realizadas usando o ambiente de computação estatística *R* (R Development Core Team, 2021).

Tabela 3.1 – Cenários considerados no estudo de simulação. Distribuição gama unitária inflacionada em zero (1-3) e distribuição gama unitária inflacionada em um (4-6).

Cenário	γ	α_0	α_1	ϕ
1	0,06	0,10	0,0	60
2	0,10	0,16	0,0	30
3	0,30	0,07	0,0	20
4	0,75	0,0	0,13	10
5	0,65	0,0	0,05	30
6	0,90	0,0	0,06	60

3.4 MEDIDAS INDIVIDUAIS

Nesta seção, apresentamos os resultados do estudo de simulação para avaliar o desempenho dos gráficos de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI) proposto por Lima-Filho et al. (2019), a partir das medidas consideradas. As Tabelas 3.2, 3.3 e 3.4 apresentam os resultados dos gráficos de controle em termos de ARL, MRL e SDRL, considerando $\alpha = 0,0027$.

Na Tabela 3.2, é possível observar que o gráfico de controle GUI_m apresenta valores de $ARL_0 = 370$, conforme o esperado quando temos um $\alpha = 0,0027$. Por outro lado, no gráfico de controle BI isso não ocorre, gerando, assim, um número mais elevado de falsos alarmes. Por exemplo, considerando o processo sob controle ($\delta = 1$) e Cenário 1, o gráfico de controle GUI_m leva em média 370 observações até a emissão de um sinal de alerta, enquanto o gráfico de controle beta inflacionado emite um sinal com, aproximadamente, 245 observações. De acordo com a Tabela 3.2 é possível verificar que à medida que o tamanho do deslocamento (δ) aumenta, os valores de ARL_1 decrescem em ambos os gráficos, isto é, os gráficos passam a detectar mais rapidamente que o processo está fora de controle.

É possível observar que a variabilidade dos dados pode causar reflexo no desempenho do gráfico de controle. Por exemplo, no Cenário 2, em que consideramos uma precisão $\phi = 30$, para $\delta = 1,25$, o gráfico de controle gama unitária inflacionado leva, em média, 87 observações para detectar uma mudança no processo. De outro modo, para o mesmo gráfico, no Cenário 1, com maior precisão ($\phi = 60$), para $\delta = 1,25$, o gráfico levaria uma média de 70 observações para identificar uma mudança de mesma magnitude. Para os cenários inflacionados em um (Tabela 3.2), observa-se que à medida

Tabela 3.3 – Performance, em termos de MRL, e limites de controle dos gráficos gama unitária inflacionado (GUI_m) e beta inflacionado (BI).

δ	Cenário 1		Cenário 2		Cenário 3	
	GUI_m	BI	GUI_m	BI	GUI_m	BI
1,00	256,12	169,44	256,12	178,18	256,12	220,62
1,05	175,29	177,53	185,60	130,33	160,15	138,66
1,10	122,78	83,41	136,73	96,90	102,14	88,89
1,15	87,86	60,45	102,28	73,14	66,39	58,08
1,20	64,11	44,67	77,60	55,99	43,94	38,65
1,25	47,64	33,60	59,65	43,42	29,60	26,17
1,30	36,00	25,69	46,43	34,09	20,28	18,03
1,35	27,63	19,94	36,55	27,06	14,12	12,62
1,40	21,52	15,70	29,09	21,72	9,99	8,98
1,45	16,98	12,53	23,39	17,60	7,18	6,49
1,50	13,58	10,12	18,98	14,40	5,23	4,75
δ	Cenário 4		Cenário 5		Cenário 6	
	GUI_m	BI	GUI_m	BI	GUI_m	BI
1,00	256,12	262,29	256,12	260,29	256,12	256,67
0,95	42,20	42,98	36,39	36,84	1,00	1,00
0,90	12,03	12,20	8,87	8,95	1,00	1,00
0,85	4,87	4,92	3,16	3,18	1,00	1,00
0,80	2,48	2,50	1,46	1,47	1,00	1,00
0,75	1,47	1,48	1,00	1,00	1,00	1,00
0,70	1,00	1,00	1,00	1,00	1,00	1,00
0,65	1,00	1,00	1,00	1,00	1,00	1,00
0,60	1,00	1,00	1,00	1,00	1,00	1,00
0,55	1,00	1,00	1,00	1,00	1,00	1,00
0,50	1,00	1,00	1,00	1,00	1,00	1,00

Tabela 3.4 – Performance, em termos de SDRL, e limites de controle dos gráficos gama unitária inflacionado (GUI_m) e beta inflacionado (BI).

δ	Cenário 1		Cenário 2		Cenário 3	
	GUI_m	BI	GUI_m	BI	GUI_m	BI
1,00	369,50	244,44	369,50	257,07	369,50	318,29
1,05	252,89	169,56	267,76	188,03	231,05	200,04
1,10	177,14	120,34	197,26	139,80	147,36	128,24
1,15	126,75	87,22	147,56	105,52	95,77	83,79
1,20	92,49	64,44	111,95	80,78	63,39	55,75
1,25	68,73	48,47	86,06	62,64	42,70	37,76
1,30	51,94	37,06	66,98	49,18	29,25	26,01
1,35	39,86	28,77	52,73	39,04	20,37	18,21
1,40	31,04	22,65	41,97	31,33	14,41	12,95
1,45	24,50	18,07	33,74	25,40	10,35	9,35
1,50	19,58	14,59	27,39	20,78	7,55	6,85
δ	Cenário 4		Cenário 5		Cenário 6	
	GUI_m	BI	GUI_m	BI	GUI_m	BI
1,00	369,50	378,41	369,50	375,52	369,50	370,29
0,95	60,88	62,00	52,49	53,15	0,95	0,95
0,90	17,36	17,61	12,79	12,91	0,24	0,24
0,85	7,02	7,10	4,55	4,58	0,22	0,22
0,80	3,56	3,59	2,08	2,10	0,22	0,22
0,75	2,10	2,12	1,13	1,13	0,21	0,21
0,70	1,38	1,38	0,68	0,68	0,20	0,20
0,65	0,96	0,97	0,43	0,43	0,19	0,19
0,60	0,71	0,71	0,29	0,29	0,19	0,19
0,55	0,54	0,54	0,20	0,20	0,18	0,18
0,50	0,42	0,43	0,15	0,15	0,17	0,17

3.5 OBSERVAÇÕES NÃO INDIVIDUAIS

A seguir são apresentados os resultados considerando subgrupos de tamanhos $s = 8, 15, 30$ e 50 . As Tabelas 3.5-3.10 apresentam as medidas de desempenho para os gráficos de controle \bar{X} , considerando cenários inflacionados em zero (1-3) e cenários inflacionados em um (4-6), para $\alpha = 0,0027$. Conforme os resultados numéricos, quando o processo está sob controle, os gráficos de controle GUI_m para observações não individuais apresentam os valores de ARL_0, MRL_0 e $SDRL_0$ próximos aos níveis nominais em todos os cenários, isto é, $ARL_0 = 370,37, MRL_0 = 256,37$ e $SDRL_0 = 369,87$ (Tabela 3.5-3.10). Por outro lado, quando o processo está sob controle, o gráfico de controle BI emite falsos alarmes com mais frequência, como por exemplo, nos Cenários 1 e 3. Conforme a perturbação na média (δ) aumenta, os gráficos de controle tendem a apresentar desempenhos próximos (Tabela 3.5 e 3.6). Por exemplo, no Cenário 4 e $\delta = 0,95$, os gráficos apresentam desempenhos similares em todos os subgrupos (Tabela 3.6).

Quando o processo está fora de controle, os valores de ARL_1, MRL_1 e $SDRL_1$ diminuem quando o tamanho do subgrupo aumenta, um reflexo da consistência do EMV. Além disso, o aumento no tamanho da amostra proporciona menor amplitude entre os limites de controle, como pode ser

observado no Cenário 1 considerando o subgrupo de tamanho $s = 8$ em relação ao subgrupo $s = 50$ (Tabela 3.5). Com o aumento no tamanho do subgrupo, os gráficos de controle apresentam menor variabilidade, detectando mais rapidamente uma mudança (de mesma magnitude) no processo. Por exemplo, no Cenário 6 com $\delta = 0,95$ e $s = 8$, considerando o gráfico de controle GUI_m , temos um valor de $SDRL_1 = 0,22$, enquanto para $s = 50$ e $\delta = 0,95$, temos $SDRL_1 = 0,01$ (Tabela 3.10).

Tabela 3.5 – Performance, em termos de ARL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
1	1,00	370,37	357,14	370,37	357,14	370,37	344,83	370,37	344,83
	1,05	192,31	192,31	200,00	200,00	83,33	83,33	50,51	50,25
	1,10	76,92	76,34	54,35	54,05	22,37	22,03	12,21	12,17
	1,15	32,89	32,79	21,88	21,88	8,15	8,03	3,96	3,94
	1,20	18,80	18,66	9,48	9,46	3,61	3,59	2,06	2,06
	1,25	10,29	10,19	5,26	5,26	2,17	2,16	1,39	1,39
	1,30	6,78	6,71	3,29	3,29	1,57	1,57	1,15	1,15
	1,35	4,47	4,44	2,39	2,39	1,29	1,29	1,05	1,05
	1,40	3,15	3,13	1,81	1,81	1,13	1,13	1,02	1,02
	1,45	2,40	2,39	1,51	1,51	1,07	1,07	1,01	1,01
	1,50	1,97	1,97	1,32	1,32	1,03	1,03	1,00	1,00
	LIC		0,0287	0,0286	0,0370	0,0369	0,0440	0,0440	0,0468
LSC		0,0917	0,0916	0,0830	0,0830	0,0757	0,0757	0,0722	0,0722
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
2	1,00	370,37	370,37	370,37	357,14	370,37	370,37	370,37	384,62
	1,05	222,22	227,27	294,12	294,12	131,58	135,14	109,89	125,00
	1,10	128,21	133,33	81,30	81,30	32,57	33,00	24,27	27,10
	1,15	59,88	60,24	32,26	32,26	12,50	12,79	7,96	8,66
	1,20	36,23	36,50	17,06	17,06	5,84	5,95	3,52	3,75
	1,25	17,61	17,76	9,41	9,41	3,36	3,40	2,07	2,15
	1,30	10,76	10,87	5,65	5,65	2,24	2,25	1,47	1,51
	1,35	7,32	7,34	3,72	3,72	1,66	1,68	1,21	1,22
	1,40	5,47	5,52	2,74	2,74	1,36	1,37	1,09	1,10
	1,45	4,00	4,03	2,17	2,17	1,22	1,22	1,04	1,04
	1,50	3,10	3,12	1,78	1,78	1,11	1,12	1,02	1,02
	LIC		0,0432	0,0432	0,0554	0,0555	0,0679	0,0679	0,0750
LSC		0,1639	0,1640	0,1470	0,1470	0,1317	0,1319	0,1254	0,1258
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
3	1,00	370,37	357,14	370,37	357,14	370,37	370,37	370,37	357,14
	1,05	227,27	227,27	62,89	62,50	45,25	45,25	25,71	25,64
	1,10	54,35	54,05	14,03	13,99	7,99	8,06	4,11	4,10
	1,15	20,28	20,08	5,08	5,05	2,68	2,69	1,65	1,64
	1,20	8,50	8,38	2,59	2,58	1,52	1,52	1,13	1,13
	1,25	4,06	4,04	1,70	1,70	1,15	1,16	1,02	1,02
	1,30	2,47	2,46	1,30	1,30	1,04	1,04	1,00	1,00
	1,35	1,78	1,78	1,14	1,14	1,01	1,01	1,00	1,00
	1,40	1,45	1,45	1,06	1,06	1,00	1,00	1,00	1,00
	1,45	1,27	1,27	1,03	1,03	1,00	1,00	1,00	1,00
	1,50	1,17	1,17	1,01	1,01	1,00	1,00	1,00	1,00
	LIC		0,1742	0,1742	0,2121	0,2121	0,2399	0,2400	0,2558
LSC		0,4036	0,4034	0,3696	0,3695	0,3521	0,3521	0,3411	0,3411

Tabela 3.6 – Performance, em termos de ARL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
4	1,00	370,37	370,37	370,37	370,37	370,37	370,37	370,37	370,37
	0,95	14,77	14,77	10,05	10,05	4,63	4,63	2,46	2,46
	0,90	3,16	3,16	1,98	1,98	1,25	1,25	1,05	1,05
	0,85	1,58	1,58	1,19	1,19	1,02	1,02	1,00	1,00
	0,80	1,19	1,19	1,04	1,04	1,00	1,00	1,00	1,00
	0,75	1,08	1,08	1,01	1,01	1,00	1,00	1,00	1,00
	0,70	1,03	1,03	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,01	1,01	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,6510	0,6510	0,6741	0,6741	0,6948	0,6948	0,7078
LSC		0,8756	0,8756	0,8376	0,8376	0,8133	0,8133	0,7962	0,7962
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
5	1,00	370,37	370,37	370,37	370,37	370,37	370,37	370,37	370,37
	0,95	11,03	11,03	4,58	4,58	2,38	2,38	1,64	1,64
	0,90	1,92	1,92	1,32	1,33	1,06	1,06	1,01	1,01
	0,85	1,21	1,21	1,05	1,05	1,00	1,00	1,00	1,00
	0,80	1,07	1,07	1,00	1,00	1,00	1,00	1,00	1,00
	0,75	1,02	1,02	1,00	1,00	1,00	1,00	1,00	1,00
	0,70	1,01	1,01	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,5808	0,5808	0,5999	0,5999	0,6132	0,6132	0,6205
LSC		0,7754	0,7754	0,7288	0,7288	0,7043	0,7043	0,6889	0,6889
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
6	1,00	370,37	357,14	370,37	370,37	370,37	370,37	370,37	370,37
	0,95	1,05	1,05	1,01	1,01	1,00	1,00	1,00	1,00
	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,85	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,8815	0,8815	0,8855	0,8855	0,8885	0,8885	0,8904
LSC		0,9375	0,9375	0,9268	0,9268	0,9168	0,9168	0,9133	0,9133

Tabela 3.7 – Performance, em termos de MRL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
1	1,00	256,37	247,21	256,37	247,21	256,37	238,67	256,37	238,67
	1,05	132,95	132,95	138,28	138,28	57,42	57,42	34,66	34,48
	1,10	52,97	52,56	37,32	37,12	15,16	14,92	8,11	8,08
	1,15	22,45	22,38	14,82	14,82	5,30	5,21	2,38	2,37
	1,20	12,68	12,58	6,22	6,20	2,14	2,12	1,05	1,04
	1,25	6,78	6,71	3,29	3,29	1,12	1,12	1,00	1,00
	1,30	4,34	4,29	1,91	1,91	1,00	1,00	1,00	1,00
	1,35	2,74	2,72	1,28	1,28	1,00	1,00	1,00	1,00
	1,40	1,81	1,80	1,00	1,00	1,00	1,00	1,00	1,00
	1,45	1,28	1,28	1,00	1,00	1,00	1,00	1,00	1,00
	1,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,0287	0,0286	0,0370	0,0369	0,0440	0,0440	0,0468
LSC		0,0917	0,0916	0,0830	0,0830	0,0757	0,0757	0,0722	0,0722
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
2	1,00	256,37	256,37	256,37	247,21	256,37	256,37	256,37	266,25
	1,05	153,69	157,19	203,52	203,52	90,86	93,32	75,82	86,30
	1,10	88,52	92,07	56,01	56,01	22,23	22,53	16,48	18,44
	1,15	41,16	41,41	22,01	22,01	8,31	8,51	5,16	5,65
	1,20	24,77	24,95	11,48	11,48	3,69	3,76	2,08	2,23
	1,25	11,85	11,96	6,17	6,17	1,96	1,99	1,05	1,11
	1,30	7,11	7,18	3,56	3,56	1,17	1,18	1,00	1,00
	1,35	4,72	4,73	2,21	2,21	1,00	1,00	1,00	1,00
	1,40	3,43	3,47	1,52	1,52	1,00	1,00	1,00	1,00
	1,45	2,41	2,43	1,12	1,12	1,00	1,00	1,00	1,00
	1,50	1,78	1,79	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,0432	0,0432	0,0554	0,0555	0,0679	0,0679	0,0750
LSC		0,1639	0,1640	0,1470	0,1470	0,1317	0,1319	0,1254	0,1258
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
3	1,00	256,37	247,21	256,37	247,21	256,37	256,37	256,37	247,21
	1,05	157,19	157,19	43,25	42,97	31,02	31,02	17,47	17,42
	1,10	37,32	37,12	9,37	9,34	5,19	5,23	2,49	2,48
	1,15	13,71	13,57	3,16	3,14	1,49	1,49	1,00	1,00
	1,20	5,54	5,46	1,42	1,42	1,00	1,00	1,00	1,00
	1,25	2,45	2,44	1,00	1,00	1,00	1,00	1,00	1,00
	1,30	1,33	1,33	1,00	1,00	1,00	1,00	1,00	1,00
	1,35	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	1,40	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	1,45	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	1,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	LIC		0,1742	0,1742	0,2121	0,2121	0,2399	0,2400	0,2558
LSC		0,4036	0,4034	0,3696	0,3695	0,3521	0,3521	0,3411	0,3411

Tabela 3.8 – Performance, em termos de MRL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
4	1,00	256,37	256,37	256,37	256,37	256,37	256,37	256,37	256,37
	0,95	9,89	9,89	6,61	6,61	2,85	2,85	1,33	1,33
	0,90	1,82	1,82	1,00	1,00	1,00	1,00	1,00	1,00
	0,85	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
LIC		0,6510	0,6510	0,6741	0,6741	0,6948	0,6948	0,7078	0,7078
LSC		0,8756	0,8756	0,8376	0,8376	0,8133	0,8133	0,7962	0,7962
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
5	1,00	256,37	256,37	256,37	256,37	256,37	256,37	256,37	256,37
	0,95	7,29	7,29	2,81	2,82	1,27	1,27	1,00	1,00
	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,85	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
LIC		0,5808	0,5808	0,5999	0,5999	0,6132	0,6132	0,6205	0,6205
LSC		0,7754	0,7754	0,7288	0,7288	0,7043	0,7043	0,6889	0,6889
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
6	1,00	256,37	247,21	256,37	256,37	256,37	256,37	256,37	256,37
	0,95	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,85	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,60	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,55	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
LIC		0,8815	0,8815	0,8855	0,8855	0,8885	0,8885	0,8904	0,8904
LSC		0,9375	0,9375	0,9268	0,9268	0,9168	0,9168	0,9133	0,9133

Tabela 3.9 – Performance, em termos de SDRL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
1	1,00	369,87	356,64	369,87	356,64	369,87	344,33	369,87	344,33
	1,05	191,81	191,81	199,50	199,50	82,83	82,83	50,00	49,75
	1,10	76,42	75,83	53,85	53,55	21,87	21,52	11,70	11,65
	1,15	32,39	32,28	21,38	21,38	7,63	7,52	3,42	3,41
	1,20	18,29	18,15	8,96	8,95	3,07	3,04	1,48	1,48
	1,25	9,78	9,68	4,74	4,73	1,60	1,58	0,74	0,74
	1,30	6,26	6,19	2,75	2,74	0,95	0,94	0,42	0,42
	1,35	3,94	3,91	1,82	1,82	0,62	0,61	0,23	0,23
	1,40	2,60	2,58	1,21	1,21	0,39	0,39	0,13	0,13
	1,45	1,83	1,82	0,88	0,88	0,27	0,26	0,07	0,07
	1,50	1,39	1,38	0,65	0,65	0,18	0,18	0,03	0,03
	LIC		0,0287	0,0286	0,0370	0,0369	0,0440	0,0440	0,0468
LSC		0,0917	0,0916	0,0830	0,0830	0,0757	0,0757	0,0722	0,0722
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
2	1,00	369,87	369,87	369,87	356,64	369,87	369,87	369,87	384,12
	1,05	221,72	226,77	293,62	293,62	131,08	134,63	109,39	124,50
	1,10	127,70	132,83	80,80	80,80	32,07	32,50	23,77	26,60
	1,15	59,38	59,74	31,75	31,75	11,99	12,28	7,45	8,14
	1,20	35,73	35,99	16,56	16,56	5,32	5,42	2,98	3,21
	1,25	17,10	17,25	8,89	8,89	2,82	2,85	1,48	1,57
	1,30	10,25	10,36	5,13	5,13	1,66	1,68	0,83	0,88
	1,35	6,80	6,82	3,18	3,18	1,05	1,06	0,50	0,52
	1,40	4,95	4,99	2,18	2,18	0,70	0,71	0,32	0,33
	1,45	3,47	3,49	1,59	1,59	0,52	0,52	0,20	0,20
	1,50	2,55	2,57	1,18	1,18	0,36	0,36	0,13	0,14
	LIC		0,0432	0,0432	0,0554	0,0555	0,0679	0,0679	0,0750
LSC		0,1639	0,1640	0,1470	0,1470	0,1317	0,1319	0,1254	0,1258
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
3	1,00	369,87	356,64	369,87	356,64	369,87	369,87	369,87	356,64
	1,05	226,77	226,77	62,39	62,00	44,75	44,75	25,20	25,14
	1,10	53,85	53,55	13,52	13,48	7,48	7,54	3,58	3,56
	1,15	19,78	19,57	4,55	4,52	2,12	2,14	1,03	1,03
	1,20	7,99	7,87	2,03	2,02	0,89	0,89	0,38	0,38
	1,25	3,53	3,50	1,09	1,09	0,42	0,42	0,15	0,15
	1,30	1,90	1,89	0,63	0,63	0,21	0,21	0,06	0,06
	1,35	1,18	1,18	0,39	0,39	0,11	0,11	0,02	0,02
	1,40	0,80	0,80	0,26	0,26	0,06	0,06	0,01	0,01
	1,45	0,58	0,58	0,16	0,16	0,02	0,02	0,01	0,01
	1,50	0,45	0,45	0,11	0,11	0,01	0,01	0,01	0,01
	LIC		0,1742	0,1742	0,2121	0,2121	0,2399	0,2400	0,2558
LSC		0,4036	0,4034	0,3696	0,3695	0,3521	0,3521	0,3411	0,3411

Tabela 3.10 – Performance, em termos de SDRL, do gráfico de controle gama unitária inflacionado (GUI_m) e beta inflacionado (BI), considerando o EMV para a média.

Cenário	δ	s = 8		s = 15		s = 30		s = 50	
		GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
4	1,00	369,87	369,87	369,87	369,87	369,87	369,87	369,87	369,87
	0,95	14,26	14,26	9,54	9,54	4,10	4,09	1,90	1,90
	0,90	2,62	2,62	1,40	1,40	0,56	0,56	0,22	0,22
	0,85	0,96	0,96	0,48	0,48	0,14	0,14	0,02	0,02
	0,80	0,48	0,48	0,20	0,20	0,03	0,03	0,01	0,01
	0,75	0,28	0,28	0,09	0,09	0,01	0,01	0,01	0,01
	0,70	0,16	0,16	0,03	0,03	0,01	0,01	0,01	0,01
	0,65	0,10	0,10	0,02	0,02	0,01	0,01	0,01	0,01
	0,60	0,07	0,07	0,01	0,01	0,01	0,01	0,01	0,01
	0,55	0,03	0,03	0,01	0,01	0,01	0,01	0,01	0,01
	0,50	0,02	0,02	0,01	0,01	0,01	0,01	0,01	0,01
	LIC		0,6510	0,6510	0,6741	0,6741	0,6948	0,6948	0,7078
LSC		0,8756	0,8756	0,8376	0,8376	0,8133	0,8133	0,7962	0,7962
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
5	1,00	369,87	369,87	369,87	369,87	369,87	369,87	369,87	369,87
	0,95	10,51	10,51	4,05	4,05	1,81	1,81	1,02	1,02
	0,90	1,33	1,33	0,66	0,66	0,25	0,25	0,08	0,08
	0,85	0,51	0,51	0,22	0,22	0,04	0,04	0,01	0,01
	0,80	0,26	0,26	0,07	0,07	0,01	0,01	0,01	0,01
	0,75	0,14	0,14	0,02	0,02	0,01	0,01	0,01	0,01
	0,70	0,08	0,08	0,01	0,01	0,01	0,01	0,01	0,01
	0,65	0,04	0,04	0,01	0,01	0,01	0,01	0,01	0,01
	0,60	0,02	0,02	0,01	0,01	0,01	0,01	0,01	0,01
	0,55	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,50	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	LIC		0,5808	0,5808	0,5999	0,5999	0,6132	0,6132	0,6205
LSC		0,7754	0,7754	0,7288	0,7288	0,7043	0,7043	0,6889	0,6889
	δ	GUI_m	BI	GUI_m	BI	GUI_m	BI	GUI_m	BI
6	1,00	369,87	356,64	369,87	369,87	369,87	369,87	369,87	369,87
	0,95	0,22	0,22	0,08	0,08	0,01	0,01	0,01	0,01
	0,90	0,03	0,03	0,01	0,01	0,01	0,01	0,01	0,01
	0,85	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,80	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,75	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,70	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,65	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,60	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,55	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	0,50	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	LIC		0,8815	0,8815	0,8855	0,8855	0,8885	0,8885	0,8904
LSC		0,9375	0,9375	0,9268	0,9268	0,9168	0,9168	0,9133	0,9133

3.6 APLICAÇÕES

Nesta seção são apresentadas duas aplicações do gráfico de controle gama unitária inflacionado a banco de dados reais. O primeiro caso consiste no monitoramento da variável taxa de letalidade por dengue na região do Caribe. A segunda aplicação apresenta o uso do gráfico de controle proposto para o monitoramento da variável umidade relativa do ar para a estação meteorológica de Araçatuba - São Paulo.

3.6.1 Dados de letalidade por dengue

Segundo a Organização Mundial de Saúde (OMS), a dengue é uma doença infecciosa febril aguda de etiologia viral que é transmitida pelo mosquito *Aedes aegypti*, sendo a mais importante arbovirose humana que afeta o ser humano e a de maior relevância nas Américas, caracterizando-se como um sério problema de saúde pública mundial. Estima-se que ocorram, anualmente, 390 milhões de infecções e 20 mil mortes no mundo (BRASIL, 2019). Uma revisão sistemática epidemiológica apontou que a incidência de dengue na região da América Latina - Caribe aumentou entre os anos de 1995-2010 (CAFFERATA et al., 2013). Entre os anos de 2001 a 2007, mais de 4 milhões de casos foram notificados nas Américas, sendo a América do Sul a subregião mais afetada, em particular, Brasil, Equador, Colômbia, Paraguai e Venezuela (SCHNEIDER, 2001).

Segundo a Pan American Health Organization (PAHO), a dengue antegiu o maior número de casos registrados nas Américas em 2019, sendo mais de 2,7 milhões (PAHO, 2019). A proporção de dengue grave (0,8%) superou a observada nos quatro anos anteriores. O número de casos notificados em 2019 a partir da Semana Epidemiológica (SE) 42 (2.733.635) é o maior registrado na história da dengue nas Américas, excedendo em 13% o número de casos notificados no ano epidêmico de 2015. Entretanto, apesar do aumento no número de casos, a taxa de letalidade foi 26% menor em 2019 (0,044%) (PAHO, 2019). Vários são os fatores para essa disseminação nos trópicos e subtropicais. O crescimento populacional, urbanização rápida e não planejada de regiões tropicais, diminuição do acesso aos cuidados de saúde e mudanças climáticas (MARTÍN et al., 2010) são alguns deles. Tendo como motivação a disseminação de casos de dengue nos últimos anos e o problema de saúde pública decorrente disso, a seguir apresentamos uma aplicação a dados de letalidade por dengue.

Nesta aplicação foram utilizados dados de letalidade por dengue para a região do Caribe Latino (Cuba, República Dominicana e Porto Rico). A Taxa de letalidade é obtida considerando o número de óbitos dentro de determinado período de tempo em relação ao número de casos registrados no mesmo intervalo. O banco de dados está disponível na Plataforma de Informação em Saúde para as Américas (PLISA), no endereço <<https://www3.paho.org/data/index.php/es/>>, para os anos de 1995 a 2022. A amostra foi dividida em dois conjuntos, sendo o primeiro com 16 observações usado para estimação dos parâmetros e limites de controle, denominada de Fase I, e o segundo com 13 observações para o monitoramento (Fase II). Os parâmetros do modelo foram estimados através do método de máxima verossimilhança. Na respectiva aplicação, consideramos a abordagem em que o tamanho do subgrupo é $s = 1$, uma vez que temos uma taxa de letalidade para cada ano. Na Tabela

3.11, observa-se que a taxa média de letalidade por dengue foi de 0,002 (desvio padrão = 0,0022) e Taxa de letalidade máxima de 0,008.

Tabela 3.11 – Medidas descritivas da variável Taxa de letalidade por dengue na região do Caribe Latino.

Mínimo	1º Quartil	Média	Mediana	3º Quartil	Máximo	Desvio-padrão	CV
0,0000	0,0000	0,0020	0,0010	0,0032	0,0080	0,0022	1,1000

Na Tabela 3.12 é possível verificar que a distribuição gama unitária inflacionada se mostra mais adequada para a modelagem dos dados de letalidade por dengue, tendo em vista os menores valores de *AIC* e *BIC*, sendo de $-137,09$ e $-134,78$, respectivamente. Na Figura 3.1 apresentamos o gráfico de controle GUI_m para o monitoramento da taxa de letalidade por dengue. Inicialmente, antes de determinarmos os limites do gráfico de controle, verificamos se $\hat{\alpha}_0(1 - \hat{\gamma}) \geq \alpha/2$. Considerando $\alpha = 0,0027$, essa relação foi satisfeita. Os valores estimados para $\hat{\gamma}$ e $\hat{\alpha}_0$ foram 0,0029 e 0,0627, respectivamente. O limite central (LC), que representa a taxa de letalidade média, foi igual a 0,0029, o limite inferior (LI) igual a zero e o limite superior (LS) sendo 0,0138. Na Figura 3.1 é possível observar que, na Fase II, não há nenhum ponto que se configure como fora de controle, ou seja, não há taxa de letalidade que esteja fora dos limites estimados.

Tabela 3.12 – Estimativas dos parâmetros e critérios de adequabilidade *AIC* e *BIC* para a variável Taxa de letalidade por Dengue.

Distribuição	Estimativa	<i>AIC</i>	<i>BIC</i>
Beta inflacionada	$\hat{\mu}$	0,0031	$-136,80$
	$\hat{\phi}$	935,636	$-134,48$
Gama unitária inflacionada	$\hat{\gamma}$	0,0029	$-137,09$
	$\hat{\phi}$	93,9270	$-134,78$

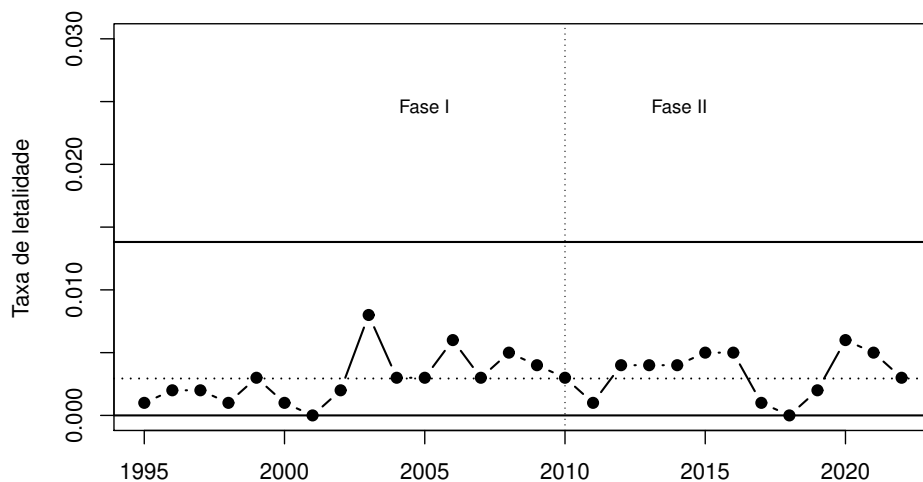


Figura 3.1 – Gráfico de controle GUI_m para a Taxa de letalidade por dengue na região do Caribe Latino, considerando $\alpha = 0,0027$. Linha tracejada referente ao limite central (LC).

3.6.2 Dados de umidade relativa

Nesta subseção apresentamos uma aplicação a dados de umidade relativa (UR). A variável UR é dada em termos percentuais, sendo resultado entre a razão da pressão parcial da água e a pressão de vapor da água em equilíbrio. A UR atinge 100% quando o ar está saturado. A umidade relativa do ar pode impactar a qualidade de vida de um indivíduo. Valores baixos de umidade relativa podem trazer tosse, desconforto na garganta ou até aparecimento de doenças no trato respiratório (YI et al., 2019). Além disso, estudos apontam que a umidade relativa alta e baixa estão associadas a um risco aumentado de influenza (WU et al., 2021).

Alguns estudos avaliaram a relação entre o COVID-19 e fatores meteorológicos como a UR (LIU et al., 2020a; YAO et al., 2020). Sarkodie e Owusu (2020) relataram que altos valores de temperatura e UR reduzem a viabilidade, estabilidade, sobrevivência e transmissão do COVID-19. Adicionalmente, Yang et al. (2021) investigaram se a relação entre fatores meteorológicos e a transmissão do COVID-19 tem ligação com especificidades sazonais e geográficas, mostrando que a temperatura e UR foram os principais fatores de transmissão, contudo, suas relações variaram com a estação e localização geográfica. Assim, monitorar a umidade relativa (UR) é de grande importância, tendo em vista as consequências que sua influência pode trazer a saúde.

Os dados utilizados nesta aplicação foram obtidos no site da Companhia Ambiental do Estado de São Paulo (CETESB), disponível em <<https://cetesb.sp.gov.br/>>. As observações foram obtidas para a estação Araçatuda, no período de 01 a 31 de julho de 2022. Nesta aplicação, consideramos o caso em que $s > 1$, uma vez que temos várias medições ao longo do dia (24 medições). Dessa forma, o dia foi considerado como o subgrupo e a umidade relativa média do dia foi modelada. A porcentagem de uns na amostra foi de 2,5%. Como é possível observar, Tabela 3.13, a média e a menor proporção de UR foram 0,6808 e 0,3400, respectivamente.

Tabela 3.13 – Medidas descritivas para a variável umidade relativa.

Mínimo	1º Quartil	Média	Mediana	3º Quartil	Máximo	Desvio-padrão	CV
0,3400	0,4900	0,6808	0,7000	0,8500	1,0000	0,1933	0,2839

De acordo com os resultados da Tabela 3.14, a distribuição gama unitária inflacionada mostra-se mais adequada para a modelagem dos dados, conforme os menores valores de AIC e BIC. Assim, para determinar os limites de controle, o conjunto de dados foi dividido em duas partes. A primeira, denominada de Fase I, composta por 360 observações, que correspondem a 15 dias de medições, foi utilizada para a estimação dos parâmetros. Usando o conjunto de dados da Fase I, os estimadores de máxima verossimilhança foram $\hat{\gamma} = 0,6849$, $\hat{\alpha}_1 = 0,0365$ e $\hat{\phi} = 1,6559$. Na sequência, considerando o tamanho da amostra por dia ($s = 24$), a distribuição empírica de $\hat{\gamma}$ foi determinada para a obtenção dos limites de controle. Para $\alpha = 0,0027$, os limites de controle inferior e superior obtidos foram 0,5658 e 0,7983, respectivamente, enquanto a linha central foi igual a 0,6844. A Figura 3.2 apresenta o gráfico de controle GUI_m no monitoramento da variável UR. É possível observar os limites de

controle estimados e a média diária de cada subgrupo. Além disso, na fase de monitoramento (Fase II), o gráfico de controle detecta um ponto fora de controle referente a amostra 26.

Tabela 3.14 – Estimativas dos parâmetros e critérios de adequabilidade AIC e BIC para a variável umidade relativa.

Distribuição		Estimativa	AIC	BIC
Beta inflacionada	$\hat{\mu}$	0,6770	-138,86	-127,20
	$\hat{\phi}$	5,1394		
Gama unitária inflacionada	$\hat{\gamma}$	0,6849	-139,32	-127,66
	$\hat{\phi}$	1,6559		

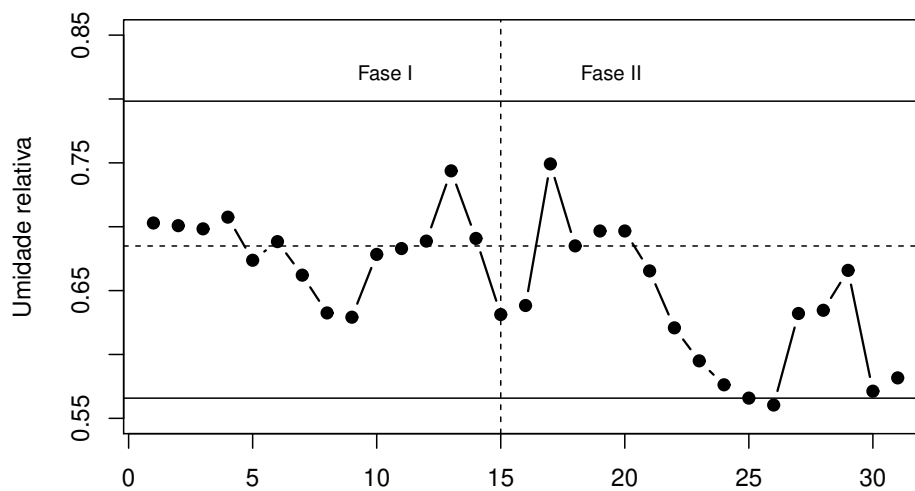


Figura 3.2 – Gráfico de controle GUI_m para a umidade relativa, considerando $\alpha = 0,0027$. Linha tracejada referente ao limite central (LC).

3.7 CONCLUSÃO

Neste artigo propomos um gráfico de controle para monitorar variáveis inflacionadas em zero ou um. O gráfico de controle proposto considera que a variável de interesse tem distribuição gama unitária inflacionada. O desempenho do gráfico proposto foi avaliado em termos da média, mediana e desvio padrão do comprimento da sequência e comparado com o gráfico de controle beta inflacionado. Os resultados numéricos indicam que o gráfico proposto apresentou bom desempenho nas abordagens consideradas. De modo geral, o gráfico de controle gama unitária inflacionado detecta de forma mais rápida uma alteração na média do processo, quando o tamanho da amostra aumenta. Esse comportamento sendo mais acentuado quando temos cenários inflacionados em um. Além disso, à medida que o deslocamento na média do processo aumenta, os gráficos de controle gama unitária inflacionado e beta inflacionado detectam de forma mais rápida uma alteração no processo. As diferenças entre os gráficos foram mais expressivas quando consideramos a abordagem com medidas individuais ($s = 1$)

e Cenários inflacionados em zero. Adicionalmente, duas aplicações a dados reais foram apresentadas para ilustrar a aplicabilidade do gráfico proposto. Levando em conta os resultados encontrados, o gráfico de controle gama unitária inflacionado pode ser bastante oportuno para monitorar variáveis dessa natureza. A partir da parametrização usada, é possível modelar a média do processo de forma a incorporar todas as informações presentes na amostra, possibilitando a obtenção de resultados mais robustos e confiáveis do ponto de vista prático. Desse modo, recomendamos o uso do gráfico de controle gama unitária inflacionado para monitorar variáveis distribuídas no intervalo unitário padrão, mas que podem assumir os valores zero ou um.

4 CONCLUSÃO GERAL

O presente estudo trouxe uma contribuição relevante na área de modelagem estatística, mais especificamente, na classe de modelos inflacionados e gráficos de controle. A distribuição gama unitária inflacionada mostra-se como uma alternativa aos modelos inflacionados (em zero, um ou zero e um) já existentes na literatura. O estudo de simulação permitiu observar o comportamento dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição e, também foi possível avaliar o comportamento de intervalos de confiança assintóticos e dos testes de hipóteses Wald, razão de verossimilhança, escore e gradiente. A aplicação a dados reais realizada a partir da distribuição, demonstra sua utilidade na modelagem de variáveis dessa natureza e reforça a ideia do uso desses modelos para solucionar problemas práticos. Além disso, a partir do gráfico de controle proposto, é possível monitorar variáveis inflacionadas em zero ou um. Tendo em vista a escassez de gráficos para monitorar variáveis dessa natureza, o uso do gráfico de controle gama unitária inflacionado torna-se bastante relevante em situações em que os gráficos de controle já existentes sejam inadequados. Vale destacar que no gráfico de controle gama unitária inflacionado, monitoramos a média da distribuição inflacionada, fazendo com que o processo de monitoramento se torne mais eficaz.

REFERÊNCIAS

- AHMAD, S.; RIAZ, M.; HUSSAIN, S.; ABBASI, S. A. On auxiliary information-based control charts for autocorrelated processes with application in manufacturing industry. **The International Journal of Advanced Manufacturing Technology**, v. 100, n. 5, p. 1965–1980, 2019.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: **Selected papers of Hirotugu Akaike**. New York: Springer, 1998. p. 199–213.
- ASLAM, M.; AZAM, M.; KHAN, N.; JUN, C. H. A control chart for an exponential distribution using multiple dependent state sampling. **Quality and Quantity**, v. 49, n. 2, p. 455–462, 2015.
- BACKER, R. C.; BROBST, R. W. Conditional double sampling. **Journal of Quality Technology**, v. 10, n. 4, p. 150–154, 1996.
- BAPAT, S. R.; BHARDWAJ, R. On an inflated Unit-Lindley Distribution. **Journal of Statistical Research**, v. 55, n. 2, p. 299–311, 2021.
- BARNDORFF-NIELSEN, O. E.; JØRGENSEN, B. Some parametric models on the simplex. **Journal of multivariate analysis**, Elsevier, v. 39, n. 1, p. 106–116, 1991.
- BAYER, F. M.; CRIBARI-NETO, F.; SANTOS, J. Inflated Kumaraswamy regressions with application to water supply and sanitation in Brazil. **Statistica Neerlandica**, Wiley Online Library, v. 75, n. 4, p. 453–481, 2021.
- BAYES, C. L.; VALDIVIESO, L. A beta inflated mean regression model for fractional response variables. **Journal of Applied Statistics**, Taylor & Francis, v. 43, n. 10, p. 1814–1830, 2016.
- BRASIL. **Guia de Vigilância em Saúde: volume único**. Brasília: Ministério da Saúde, 2019. Disponível em: <http://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_saude_3ed.pdf>. Acesso em: 25 mai. 2023.
- BUSE, A. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. **The American Statistician**, Taylor & Francis, v. 36, n. 3a, p. 153–157, 1982.
- CAFFERATA, M. L.; BARDACH, A.; REY-ARES, L.; ALCARAZ, A.; CORMICK, G.; GIBBONS, L.; ROMANO, M.; CESARONI, S.; RUVINSKY, S. Dengue epidemiology and burden of disease in Latin America and the Caribbean: a systematic review of the literature and meta-analysis. **Value in health regional issues**, Elsevier, v. 2, n. 3, p. 347–356, 2013.
- CASELLA, G.; BERGER, R. **Statistical Inference**. 2nd: Duxbury Press, Pacific Grove, 2002.
- CASTAGLIOLA, P.; MARAVELAKIS, P. E. A CUSUM control chart for monitoring the variance when parameters are estimated. **Journal of Statistical Planning and Inference**, Elsevier, v. 141, n. 4, p. 1463–1478, 2011.
- CORPES, G. do N. **Distribuição beta prime inflacionada em zero**. 66 p. Dissertação (Mestrado) — Universidade Federal da Paraíba, João Pessoa, 2021.
- COX, D. R.; HINKLEY, D. V. **Theoretical Statistics**. New York: CRC Press, 1979.
- CRIBARI-NETO, F.; SANTOS, J. Inflated kumaraswamy distributions. **Annals of the Brazilian Academy of Sciences**, v. 91, n. 2, 2019.

- DEY, S.; MENEZES, A. F.; MAZUCHELI, J. Comparison of estimation methods for unit-gamma distribution. **Journal of data science**, v. 17, n. 4, p. 768–801, 2019.
- DUARTE, B. P. M.; SARAIVA, P. M. An optimization-based approach for designing attribute acceptance sampling plan. **International Journal of Quality and Reliability**, v. 25, n. 8, p. 824–841, 2008.
- ESPINHEIRA, P. L.; SILVA, A. de O. Residual and influence analysis to a general class of simplex regression. **Test**, v. 29, n. 2, p. 523–552, 2020.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004.
- FERREIRA, A. B.; MAZUCHELI, J. The zero, one and zero-and-one-inflated new Unit-Lindley distributions. **Brazilian Journal of Biometrics**, v. 40, n. 3, 2022.
- GOVE, A. D.; SALDER, R.; MATSUKI, M.; ARCHIBALD, R.; PEARSE, S.; GARKAKLIS, M. Control charts for improved decisions in environmental management: A case study of catchment water supply in south-west western australia. **Ecological Management and Restoration**, v. 14, n. 2, p. 127–134, 2013.
- GRASSIA, A. On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions. **Australian Journal of Statistics**, v. 19, n. 2, p. 108–114, 1977.
- GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. Modified likelihood ratio tests for unit gamma regressions. **Journal of Applied Statistics**, Taylor & Francis, v. 47, n. 9, p. 1562–1586, 2019.
- GUPTA, A. K.; NADARAJAH, S. **Handbook of beta distribution and its applications**. USA: CRC press, 2004.
- HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 41, n. 2, p. 190–195, 1979.
- HO, L. L.; FERNANDES, F. H.; BOURGUIGNON, M. Control charts to monitor rates and proportions. **Quality and Reliability Engineering International**, v. 35, n. 1, p. 74–83, 2019.
- Instituto Trata Brasil. **Ranking do Saneamento**. São Paulo, 2022.
- JOEKES, S.; SMREKAR, M.; BARBOSA, E. P. Extending a double sampling control chart for non-conforming proportion in high quality processes to the case of small samples. **Statistical Methodology**, Elsevier, v. 23, p. 35–49, 2015.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions**. New York: John Wiley & Sons, 1995. v. 2.
- JORGE, N.; STEPHEN, J. W. **Numerical optimization**. New York, NY: Springer, 2006.
- KHALIQ, Q.; RIAZ, M.; AHMAD, S. On designing a new Tukey-EWMA control chart for process monitoring. **The International Journal of Advanced Manufacturing Technology**, Springer, v. 82, p. 1–23, 2016.
- KUMARASWAMY, P. Sinpower probability density function. **Journal of Hydrology**, Elsevier, v. 31, n. 1-2, p. 181–184, 1976.

- KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, v. 46, n. 1-2, p. 79–88, 1980.
- LEHMANN, E. L.; ROMANO, J. P. **Testing Statistical Hypotheses**. New York: Springer, 1986.
- LEIVA, V.; MARCHANT, C.; RUGGERI, F.; SAULO, H. A criterion for environmental assessment using birnbaum-saunders attribute control charts. **Environmetrics**, v. 26, n. 7, p. 463–476, 2015.
- LEMONTE, A. **The Gradient Test: Another Likelihood-Based Test**. New York: Academic Press, 2016.
- LIMA-FILHO, L. M. A.; BAYER, F. M. Kumaraswamy control chart for monitoring double bounded environmental data. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 50, n. 9, p. 2513–2528, 2021.
- LIMA-FILHO, L. M. A.; BOURGUIGNON, M.; HO, L.; FERNANDES, F. H. L. Median control charts for monitoring asymmetric quality characteristics double bounded. **Quality and Reliability Engineering International**, v. 36, n. 7, p. 2285–2308, 2020a.
- LIMA-FILHO, L. M. A.; PEREIRA, T. L.; BAYER, F. M. Inflated beta control chart for monitoring double bounded processes. **Computers and Industrial Engineering**, v. 136, p. 265–276, 2019.
- LIMA-FILHO, L. M. A.; PEREIRA, T. L.; BAYER, F. M.; SOUZA, T. C. de; BOURGUIGNON, M. Control chart for monitoring zero-or-one inflated double-bounded environmental processes. **Environmental and Ecological Statistics**, Springer, v. 30, n. 3, p. 355–377, 2023.
- LIN, C. Z.; MUKHERJEE, A.; CHONG, M. B. Some simplified Shewhart-type distribution-free joint monitoring schemes and its application in monitoring drinking water turbidity. **Quality Engineering**, v. 32, p. 91–110, 2020.
- LIU, J.; ZHOU, J.; YAO, J.; ZHANG, X.; LI, L.; XU, X.; HE, X.; WANG, B.; FU, S.; NIU, T. et al. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. **Science of the total environment**, Elsevier, v. 726, p. 138513, 2020.
- LIU, P.; YUEN, K. C.; WU, L.-C.; TIAN, G.-L.; LI, T. Zero-one-inflated simplex regression models for the analysis of continuous proportion data. **Statistics and Its Interface**, International Press of Boston, v. 13, n. 2, p. 193–208, 2020.
- LOMBARD, F.; MAXWELL, R. A CUSUM procedure to detect deviations from uniformity in angular data. **Journal of Applied Statistics**, v. 39, p. 1871–1880, 2012.
- LUND, R.; SEYMOUR, L. Assessing temperature anomalies for a geographical region: a control chart approach. **Environmetrics: The Official Journal of the International Environmetrics Society**, Wiley Online Library, v. 10, n. 2, p. 163–177, 1999.
- MARTÍN, J. L. S.; BRATHWAITE, O.; ZAMBRANO, B.; SOLÓRZANO, J. O.; BOUCKENOOGHE, A.; DAYAN, G. H.; GUZMÁN, M. G. The epidemiology of dengue in the Americas over the last three decades: a worrisome reality. **The American journal of tropical medicine and hygiene**, The American Society of Tropical Medicine and Hygiene, v. 82, n. 1, p. 128, 2010.
- MARTINEZ-FLOREZ, G.; OLMOS, N. M.; VENEGAS, O. Unit-bimodal Birnbaum-Saunders distribution with applications. **Communications in Statistics - Simulation and Computation**, p. 1–20, 2022a.
- MARTINEZ-FLOREZ, G.; TOVAR-FALÓN, R.; BARRERA-CAUSIL, C. Inflated Unit-Birnbaum-Saunders Distribution. **Mathematics**, v. 10, n. 4, p. 667, 2022b.

- MAZUCHELI, J.; MENEZES, A.; GHITANY, M. The Unit-Weibull distribution and associated inference. **Journal of Applied Probability and Statistics**, v. 13, n. 2, p. 1–22, 2018.
- MAZUCHELI, J.; MENEZES, A. F. B.; DEY, S. Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution. **Communications in Statistics - Theory and Methods**, v. 47, n. 15, p. 3767–3778, 2018.
- MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. **Econometric foundations pack with CD-ROM**. [S.l.]: Cambridge University Press, 2000.
- MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. 6th Edition. ed. New York: John Wiley & Sons, 2009.
- MOUSA, A. M.; EL-SHEIKH, A. A.; ABDEL-FATTAH, M. A. A gamma regression for bounded continuous variables. **Advances and Applications in Statistics**, v. 49, n. 4, p. 305–326, 2016.
- OSPINA, R.; FERRARI, S. L. A general class of zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1609–1623, 2012.
- OSPINA, R.; FERRARI, S. L. P. Inflated beta distributions. **Statistical Papers**, v. 51, p. 111–126, 2010.
- PAHO. **Epidemiological Update: Dengue**. Washington, D.C., 2019. Disponível em: <https://www3.paho.org/hq/index.php?option=com_docman&view=download&category_slug=dengue-2217&alias=50963-11-november-2019-dengue-epidemiological-update-1&Itemid=270&lang=en>. Acesso em: 28 mai. 2023.
- PAROISSIN, C.; PENALVA, L.; PÉTRAU, A.; VERDIER, G. New control chart for monitoring and classification of environmental data. **Environmetrics**, v. 27, n. 3, p. 182–193, 2016.
- PAWITAN, Y. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.
- Pesquisa Nacional por Amostra de Domicílios Contínua. **Características gerais dos domicílios e dos moradores 2019**. Rio de Janeiro, Brasil, 2020.
- PRESS, W. H. **Numerical Recipes 3rd edition: The Art of Scientific Computing**. New York: Cambridge University Press, 2007.
- PUMI, G.; RAUBER, C.; BAYER, F. M. Kumaraswamy regression model with Aranda-Ordaz link function. **Test**, v. 29, n. 4, p. 1051–1071, 2020.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Acesso em: 11 jun. 2023.
- RAHMAN, T.; HAZARIKA, P. J.; ALI, M. M.; BARMAN, M. P. Three-Inflated Poisson Distribution and its Application in Suicide Cases of India During Covid-19 Pandemic. **Annals of Data Science**, v. 9, p. 1103–1127, 2022.
- RATNAPARKHL, M. V.; MOSIMANN, J. E. On the normality of transformed beta and unit-gamma random variables. **Communications in Statistics-Theory and Methods**, v. 19, n. 10, p. 3833–3854, 1990.
- RIAZ, M. A sensitive non-parametric EWMA control chart. **Journal of the Chinese Institute of Engineers**, v. 38, n. 2, p. 208–219, 2015.

- ROCHA, S. S.; ESPINHEIRA, P. L.; CRIBARI-NETO, F. Residual and local influence analyses for unit gamma regressions. **Statistica Neerlandica**, Wiley Online Library, v. 75, n. 2, p. 137–160, 2021.
- SAGRILLO, M.; GUERRA, R. R.; BAYER, F. M. Modified Kumaraswamy distributions for double bounded hydro-environmental data. **Journal of Hydrology**, v. 603, p. 127021, 2021.
- SANT'ANNA, A. M. O.; CATEN, C. S. ten. Beta control charts for monitoring fraction data. **Expert Systems with Applications**, v. 39, n. 11, p. 10236–10243, 2012.
- SARKODIE, S. A.; OWUSU, P. A. Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. **Environmental Research**, Elsevier, v. 191, p. 110101, 2020.
- SCHNEIDER, J. A timeline for dengue in the Americas to December 31, 2000 and noted first occurrences. **The History of Dengue and Dengue Hemorrhagic Fever (DHF) in the Region of the Americas, 1635-2001**, Pan American Health Office, 2001.
- SCHUWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978.
- Secretaria Nacional de Saneamento - SNS. **Diagnóstico Temático Serviços de Água e Esgoto**. Brasília, 2022.
- SEN, P. K.; SINGER, J. M. **Large sample methods in statistics: an introduction with applications**. New York: CRC press, 1994.
- SUGIURA, N. Further analysis of the data by Akaike's information criterion and the finite corrections: further analysis of the data by Akaike's. **Communications in Statistics-theory and Methods**, Taylor & Francis, v. 7, n. 1, p. 13–26, 1978.
- TEOH, W. L.; CHONG, J. K.; KHOO M. B. C. CASTAGLIOLA, P.; YEONG, W. C. Optimal designs of the variable sample size \bar{X} chart based on median run length and expected median run length. **Quality and Reliability Engineering International**, v. 33, n. 1, p. 121–134, 2017.
- TERRELL, G. R. The Gradient Statistic. **Computing Science and Statistics**, v. 34, p. 206–215, 2002.
- TU, W. Zero-inflated data. **Encyclopedia of environmetrics**, Wiley Online Library, 2006.
- VRIES, A. D.; RENEAU, J. K. Application of statistical process control charts to monitor changes in animal production systems. **Journal of Animal Science**, v. 88, p. 11–24, 2010.
- WELSH, A. H. **Aspects of Statistical Inference**. New York: John Wiley & Sons, 2011.
- WU, Q.; HE, J.; ZHANG, W.-Y.; ZHAO, K.-F.; JIN, J.; YU, J.-L.; CHEN, Q.-Q.; HOU, S.; ZHU, M.; XU, Z. et al. The contrasting relationships of relative humidity with influenza A and B in a humid subtropical region. **Environmental Science and Pollution Research**, Springer, v. 28, p. 36828–36836, 2021.
- YANG, X.-D.; LI, H.-L.; CAO, Y.-E. Influence of meteorological factors on the COVID-19 transmission with season and geographic location. **International Journal of Environmental Research and Public Health**, MDPI, v. 18, n. 2, p. 484, 2021.
- YAO, M.; ZHANG, L.; MA, J.; ZHOU, L. On airborne transmission and control of SARS-CoV-2. **Science of The Total Environment**, Elsevier, v. 731, p. 139178, 2020.

YI, L.; XU, X.; GE, W.; XUE, H.; LI, J.; LI, D.; WANG, C.; WU, H.; LIU, X.; ZHENG, D. et al. The impact of climate variability on infectious disease transmission in China: Current knowledge and further directions. **Environmental research**, Elsevier, v. 173, p. 255–261, 2019.

YOO, S.-H. A note on an approximation of the mobile communications expenditures distribution function using a mixture model. **Journal of Applied Statistics**, v. 31, n. 7, p. 747–752, 2004.

APÊNDICE A –

A seguir são apresentadas as expressões das primeiras derivadas com relação a cada um dos parâmetros do modelo.

$$U(\alpha_0) = \frac{\partial \ell(\theta)}{\partial \alpha_0} = \frac{\sum_{i=1}^n \mathbb{1}_{\{0\}}(y_i)}{\alpha_0} - \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{1 - \gamma}{c} \\ + \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{d(1+d)}{\check{\mu}\phi} (y_i^* - \mu^*) \frac{\gamma(1-\gamma)(1-\alpha_1)}{c^2},$$

$$U(\alpha_1) = \frac{\partial \ell(\theta)}{\partial \alpha_1} = \frac{\sum_{i=1}^n \mathbb{1}_{\{1\}}(y_i)}{\alpha_1} - \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{\gamma}{c} \\ - \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{d(1+d)}{\check{\mu}\phi} (y_i^* - \mu^*) \frac{\gamma(1-\gamma)(1-\alpha_0)}{c^2},$$

$$U(\gamma) = \frac{\partial \ell(\theta)}{\partial \gamma} = - \sum_{i=1}^n \mathbb{1}_{\{0\}}(y_i) \frac{1}{1-\gamma} + \sum_{i=1}^n \mathbb{1}_{\{1\}}(y_i) \frac{1}{\gamma} + \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{\alpha_0 - \alpha_1}{c} \\ + \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \frac{d(1+d)}{\check{\mu}\phi} (y_i^* - \mu^*) \frac{(1-\alpha_0)(1-\alpha_1)}{c^2},$$

$$U(\phi) = \frac{\partial \ell(\theta)}{\partial \phi} = \sum_{i=1}^n (1 - \mathbb{1}_{\{0\}}(y_i))(1 - \mathbb{1}_{\{1\}}(y_i)) \left\{ \log(-\log(y_i)) - \frac{d \log(\check{\mu})}{\phi} \left[1 + \frac{d \log(y_i)}{\phi \check{\mu}^{1/\phi}} \right] \right. \\ \left. - \log\left(\frac{\check{\mu}^{1/\phi}}{d}\right) - \psi(\phi) \right\}.$$

As expressões para os elementos que compõem a matriz de informação de Fisher apresentada em (9), são dadas por

$$K_{\alpha_0 \alpha_0} = \frac{n(1-\gamma)}{\alpha_0} + \frac{n(1-\gamma)^2}{c} + \frac{n(1+d)^2 [\gamma(1-\gamma)(1-\alpha_1)]^2}{\check{\mu}^2 \phi c^3},$$

$$K_{\alpha_1 \alpha_1} = \frac{n\gamma}{\alpha_1} + \frac{n\gamma^2}{c} + \frac{n(1+d)^2 [\gamma(1-\gamma)(1-\alpha_0)]^2}{\check{\mu}^2 \phi c^3},$$

$$K_{\gamma\gamma} = \frac{n\alpha_0}{1-\gamma} + \frac{n\alpha_1}{\gamma} + \frac{n(\alpha_0 - \alpha_1)^2}{c} + \frac{n(1+d)^2 [(1-\alpha_0)(1-\alpha_1)]^2}{\check{\mu}^2\phi c^3},$$

$$K_{\phi\phi} = n(1 - \alpha_0(1-\gamma) - \alpha_1\gamma) \left\{ \frac{2d \log(\check{\mu})}{\phi^2 \check{\mu}^{1/\phi}} + \frac{d^2 (\log(\check{\mu}))^2}{\phi^3 \check{\mu}^{2/\phi}} + \psi'(\phi) \right\},$$

$$K_{\alpha_0\alpha_1} = K_{\alpha_1\alpha_0} = -\frac{n\gamma(1-\gamma)}{c} - \frac{n(1+d)^2}{\check{\mu}^2\phi} \left(\frac{\gamma^2(1-\gamma)^2(1-\alpha_0)(1-\alpha_1)}{c^3} \right),$$

$$K_{\alpha_0\gamma} = K_{\gamma\alpha_0} = -n - \frac{n(1-\gamma)(\alpha_0 - \alpha_1)}{c} + \frac{n(1+d)^2}{\check{\mu}^2\phi} \left(\frac{\gamma(1-\gamma)(1-\alpha_0)(1-\alpha_1)^2}{c^3} \right),$$

$$K_{\alpha_0\phi} = K_{\phi\alpha_0} = -\frac{n\gamma(1-\gamma)(1-\alpha_1)}{c} \left[\frac{d(1+d)}{\check{\mu}\phi} \left(\frac{1}{d} + \frac{\log(\check{\mu})}{\phi\check{\mu}^{1/\phi}} \right) \right],$$

$$K_{\alpha_1\gamma} = K_{\gamma\alpha_1} = n - \frac{n\gamma(\alpha_0 - \alpha_1)}{c} - \frac{n(1+d)^2}{\check{\mu}^2\phi} \left(\frac{\gamma(1-\gamma)(1-\alpha_1)(1-\alpha_0)^2}{c^3} \right),$$

$$K_{\alpha_1\phi} = K_{\phi\alpha_1} = \frac{n\gamma(1-\gamma)(1-\alpha_0)}{c} \left[\frac{d(1+d)}{\check{\mu}\phi} \left(\frac{1}{d} + \frac{\log(\check{\mu})}{\phi\check{\mu}^{1/\phi}} \right) \right],$$

$$K_{\gamma\phi} = K_{\phi\gamma} = -n(1 - \alpha_0(1-\gamma) - \alpha_1\gamma) \left(\frac{(1-\alpha_0)(1-\alpha_1)}{c^2} \right) \left[\frac{d(1+d)}{\check{\mu}\phi} \left(\frac{1}{d} + \frac{\log(\check{\mu})}{\phi\check{\mu}^{1/\phi}} \right) \right].$$