
Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Gráfico de controle CUSUM para monitorar a média
de processos limitados usando variável auxiliar

Letícia Garcez Corrêa da Costa

Dezembro/2021

Letícia Garcez Corrêa da Costa

Gráfico de controle CUSUM para monitorar a média de processos limitados usando variável auxiliar

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Estatística Aplicada.

João Pessoa
Dezembro de 2021

Catálogo na publicação
Seção de Catalogação e Classificação

C837g Costa, Leticia Garcez Corrêa da.

Gráfico de controle CUSUM para monitorar a média de processos limitados usando variável auxiliar / Leticia Garcez Corrêa da Costa. - João Pessoa, 2021.

40 p. : il.

Orientação: Luiz Medeiros de Araújo Lima Filho.

TCC (Graduação/Bacharelado em Estatística) -
UFPB/CCEN.

1. Gráfico de controle. 2. CUSUM. 3. Distribuição beta inflacionada. 4. Modelo de regressão. I. Lima Filho, Luiz Medeiros de Araújo. II. Título.

UFPB/CCEN

CDU 519.171(043.2)



ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

“Gráfico de controle CUSUM para monitorar a média de processos limitados usando variável auxiliar”

Leticia Garcez Corrêa da Costa

No primeiro dia do mês de Dezembro de 2021 às 10:00 horas, de modo remoto, via plataforma Google Meet, através de videoconferência no endereço eletrônico: <https://meet.google.com/eaq-ooay-hhp>, constituiu-se a Banca Examinadora do Trabalho de Conclusão de Curso da discente Leticia Garcez Corrêa da Costa, matrícula nº 20170034555, composta pelos professores: Dr. Luiz Medeiros de Araujo Lima Filho, Presidente/Orientador (Departamento de Estatística - UFPB), Dr. Eufrásio de Andrade Lima Neto, Examinador (Departamento de Estatística - UFPB), Dra. Tarciana Liberal Pereira de Araujo, Examinadora (Departamento de Estatística - UFPB) e Dra. Tatiene Correia de Souza, Examinadora Suplente (Departamento de Estatística - UFPB). Iniciando-se os trabalhos, o presidente da Banca Examinadora cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra à candidata para que se fizesse, oralmente, a exposição do Trabalho de Conclusão de Curso intitulado “Gráfico de controle CUSUM para monitorar a média de processos limitados usando variável auxiliar”. Concluída a apresentação, a candidata foi arguida pela Banca Examinadora que sugeriu que a discente fizesse algumas alterações até o dia 15 de Dezembro de 2021. Uma vez entregue a versão final do Trabalho de Conclusão de Curso à Coordenação do Bacharelado em Estatística com as alterações solicitadas pela Banca Examinadora, dentro do prazo do prazo estabelecido, a discente será aprovada com a nota 10 (DEZ), resultante da média aritmética das notas atribuídas pelos membros da Banca Examinadora.

Dr. Luiz Medeiros de Araujo Lima Filho
(Professor Orientador)

Dr. Eufrásio de Andrade Lima Neto
(Professor Examinador)

Dra. Tarciana Liberal Pereira de Araujo
(Professora Examinadora)

Leticia Garcez Corrêa da Costa
(Discente)

João Pessoa, 01 de Dezembro de 2021.

Dedico esse estudo à todos que estiveram ao meu lado durante a graduação, em especial minha família.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Gustavo e Rosane, que me motivaram durante todo o processo da graduação e me deram todo o apoio que eu precisava para morar em uma cidade nova. Foram essenciais para as decisões que tomei durante a estadia em João Pessoa, sempre me orientando para as melhores escolhas.

A minha irmã Priscilla que sempre esteve ao meu lado mesmo distante, me visitando quando possível e compartilhando bons momentos. Obrigada por apoiar minhas decisões, sempre demonstrar o quão orgulhosa fica com minhas conquistas pessoais e profissionais.

A família que me acolheu no período da graduação, em especial a Maria Inês que me deu apoio e me ajudou em tudo que eu precisava. Agradeço imensamente por terem me feito sentir parte da família.

Ao meu namorado Nicollas Gabriel, também colega de curso, que está comigo desde o segundo semestre da faculdade. Agradeço pela paciência, apoio e todo o incentivo que me dá. Estar ao seu lado foi fundamental para meu crescimento profissional e pessoal. O período longe de casa não teria sido a mesma coisa sem a sua parceria, obrigada por tudo.

Aos meus colegas de graduação, em especial Wanusa, Bianca, Carol e Cecília. Obrigada por compartilharem comigo esses 4 anos. Todos os momentos de dificuldade que enfrentamos em algumas cadeiras do curso sem dúvida foram mais leves com vocês por perto.

Também vale agradecer aos colegas do Hospital Nossa Senhora das Neves, onde passei mais de um ano, em especial ao Adriano. Obrigada por todos os ensinamentos e pelo tempo tentando aprender novos assuntos, saiba que sou muito grata pela sua paciência e amizade.

Aos professores do Departamento de Estatística que sempre se propuseram a ajudar e tirar todas as dúvidas que vinham a surgir. Obrigada pelo profissionalismo e pela dedicação com o ensino. Em especial ao professor Luiz, que pude ter contato durante boa parte do curso, desde a orientação que tive na tutoria de Estatística Vital em 2018,

PIBIC em 2019 e 2020 e dando continuidade em 2021 com a última etapa da graduação, o trabalho de conclusão de curso. Muito obrigada, sua paciência foi essencial para meu crescimento profissional.

A todos meus familiares que estive longe durante esses 4 anos em João Pessoa, em especial minha avó Maria Cecília, meus tios Neto, Dariella e Anna Karolina, meus primos Felipe, João Pedro e Thiago. Senti muita falta de todos, mas sempre entenderam minha posição e a decisão de ficar longe de casa esses anos, obrigada por tudo!

*“A verdadeira motivação vem de realização,
desenvolvimento pessoal, satisfação
no trabalho e reconhecimento.”*

(Frederick Herzberg)

O Controle Estatístico de Processos (CEP) busca o controle da qualidade do produto durante todo o processo produtivo, procurando dar estabilidade ao processo e reduzir a variabilidade do processo. Uma poderosa ferramenta para a análise e monitoramento da qualidade é o gráfico de controle. Gráficos de controle tradicionais como Shewhart são muito úteis quando o intuito é detectar grandes desvios na média do processo. Porém, esse tipo de gráfico perde eficiência quando o objetivo é detectar pequenas variações. Uma alternativa encontrada é utilizar o gráfico de controle da soma cumulativa (CUSUM). Esse tipo de gráfico armazena informações acumuladas de toda a sequência de pontos, tornando-o mais sensível para detectar pequenas mudanças. Em algumas situações a característica da qualidade pode ser correlacionada com uma variável de controle, também chamada de variável auxiliar. Nesses casos, o gráfico de controle baseado em regressão pode ser uma ferramenta de controle eficaz. O presente estudo propõe o gráfico de controle CUSUM baseado no resíduo de Pearson do modelo de regressão beta inflacionado em zero ou um para monitorar variáveis limitadas na presença de variáveis auxiliares. O gráfico de controle proposto é avaliado e comparado com o gráfico de controle de regressão beta inflacionado do tipo Shewhart com o objetivo de investigar a sensibilidade quando o processo encontra-se em controle e fora de controle. O desempenho dos gráficos de controle são avaliados e comparados por meio de um extenso estudo de simulação de Monte Carlo através do comprimento da sequência (*run lenght* - RL). Todos os códigos são implementados por meio do *software* R. Por fim, para demonstrar a aplicabilidade do gráfico proposto, uma aplicação a dados reais é considerada.

Palavras-chave: Gráfico de controle; CUSUM; Distribuição beta inflacionada; Modelo de regressão.

Statistical Process Control (SPC) searches for the product control of quality all over the productive process, looking for process stability and reducing the process variability. A powerful tool for analyzing and monitoring the quality is the control chart. Traditional control charts like the Shewhart type are really used to detect big shifts at the mean process. However, this type of chart loses efficiency when the goal is to detect small shifts. An alternative is to use the cumulative sum control chart (CUSUM). This kind of chart receives cumulative information of the entire sample sequence, making it more sensitive to detect small shifts. The CUSUM must be the most suitable control chart to use in these cases. In some situations the quality characteristic is correlated with one or more control variables, also called control auxiliary. In these cases the regression control chart can be an effective control tool. This study proposes a CUSUM control chart based on Pearson residuals of an inflated zero or one beta regression model to monitor double bounded processes with the presence of control variables. The proposed control chart is evaluated and compared with an inflated beta regression control chart to investigate the sensibility when the process is in control and out of control. An extensive Monte Carlo study is conducted to compare the performance of both control charts through the run length (RL). All codes are implemented through the R software. Finally, real data are used to show the applicability of the proposed control chart.

Keywords: Control chart; CUSUM; Inflated beta distribution; Regression model.

Lista de Figuras	viii
Lista de Tabelas	ix
1 Introdução	1
1.1 Organização do Trabalho	3
2 Regressão Beta Inflacionada	5
2.1 Modelo de Regressão	6
2.2 Resíduo	7
3 Gráficos de Controle	8
3.1 Gráfico de controle para regressão	10
3.1.1 Gráfico de controle de regressão beta inflacionado	10
3.2 CUSUM baseado no resíduo de Pearson do modelo de regressão beta inflacionado	13
4 Simulações	15
4.1 Estudo de Simulação	15
4.1.1 Desempenho dos gráficos de controle	17
4.1.2 Algoritmo	18
4.2 Análise de Sensibilidade	19
5 Aplicações	23
6 Considerações finais	27
Referências bibliográficas	29
A Resultados da Simulação	31

LISTA DE FIGURAS

3.1	Gráfico de controle.	9
3.2	Gráfico de controle de regressão.	10
3.3	Gráfico da soma cumulativa.	13
4.1	Valores de \widehat{ARL} para os cenários 1, 2 e 3 utilizando CUSUM-IBRCC (linha tracejada azul) e IBRCC (linha contínua vermelha)	21
4.2	Valores de \widehat{ARL} para os cenários 4, 5 e 6 utilizando CUSUM-IBRCC (linha tracejada azul) e IBRCC (linha contínua vermelha)	22
5.1	Histograma da umidade relativa na cidade Albury - Austrália nos meses de agosto de 2009 a 2019	24
5.2	Gráfico de controle CUSUM-IBRCC	25
5.3	Gráfico de controle IBRCC	26

LISTA DE TABELAS

4.1	Cenários considerados no estudo de simulação	15
4.2	Característica dos cenários	16
5.1	Descrição das variáveis utilizadas no modelo de regressão	23
5.2	Estatística descritiva das variáveis utilizadas no modelo	24
5.3	Modelo de regressão ajustado para os dados de umidade relativa	25
A.1	Comparação da performance dos gráficos de controle inflacionados em termos de ARL, MRL e SDRL.	32

CAPÍTULO 1

INTRODUÇÃO

Atualmente, há diversas empresas no mercado procurando seu espaço. Para que a empresa se destaque, tem que seguir cada vez mais um rigoroso padrão de qualidade de seus produtos. A qualidade sempre foi uma característica relevante para todos produtos e serviços, porém, sua importância tem se intensificado com o decorrer do tempo. A qualidade de um produto ou serviço tem sido o fator chave na decisão dos clientes. Sendo assim, a qualidade é considerada um dos principais pilares para o sucesso de uma organização. Para isso, houve necessidade da criação de técnicas estatísticas para controlar o processo e buscar melhorias.

O Controle Estatístico de Processos (CEP) é um método para monitorar produtos durante todo seu desenvolvimento, procurando identificar itens que não seguem o padrão desejado. É uma forma de prevenção e detecção de possíveis defeitos que possam vir a aparecer nos produtos e desempenha um papel fundamental principalmente nas indústrias. Além disso, é uma forma estatística de minimizar a variabilidade e garantir a qualidade.

A qualidade sempre foi uma característica muito importante para qualquer produto ou serviço. Para isso, métodos estatísticos têm papel primordial para a melhoria da qualidade. A relevância do controle estatístico da qualidade se intensificou principalmente após a Segunda Guerra Mundial, em que ficou evidente a necessidade e a eficiência do controle estatístico para acompanhar e melhorar a qualidade do produto. Era necessário produzir grandes quantidades de produtos militares com alta qualidade e em um pequeno espaço de tempo. Para que o nível de qualidade desejado fosse atingido fez-se necessária a utilização de diversas técnicas estatísticas para a melhoria contínua do processo. Houve uma grande expansão dessas técnicas principalmente nas indústrias de manufatura, onde inicialmente era o principal ramo que utilizava a ferramenta.

De acordo com Montgomery (2009), o CEP pode ser aplicado a qualquer processo, tendo sete principais ferramentas, em que uma delas é o gráfico de controle. Essa ferramenta é um meio visual de identificar se o processo necessita de ações corretivas, melhorando o processo e acompanhando a variabilidade.

O pioneiro na aplicação de técnicas estatísticas para o controle da qualidade foi Walter Andrew Shewart da companhia telefônica *Bell Telephone Laboratories*, que desenvolveu o gráfico de controle nos Estados Unidos da América em 1924. Essa etapa da história ficou conhecida como o começo formal do controle estatístico da qualidade. Segundo Montgomery (2009), os gráficos de controle monitoram a estabilidade do processo, fornecem informações úteis para a melhoria, estimam os parâmetros de um processo de produção e assim determinam a capacidade do processo.

Há duas formas de definir os limites de controle, sendo o caso K e o caso U. No caso K os parâmetros são conhecidos, enquanto no caso U eles são desconhecidos. Em muitas situações práticas, os parâmetros são desconhecidos e algum método de estimação se faz necessário. Esse processo será feito na etapa conhecida como Fase I. Já na Fase II, ocorre o monitoramento do processo, verificando se está sob controle ou fora de controle. O sucesso da Fase I afeta a performance do monitoramento da qualidade na Fase II (CHAKRABORTI; HUMAN; GRAHAM, 2008).

Dito isso, os gráficos de controle têm sido a principal ferramenta para avaliar e controlar os processos. Há situações em que o interesse é modelar variáveis limitadas. Exemplos dessas variáveis podem ser dados de taxas e proporções (HO; FERNANDES; BOURGUIGNON, 2019). Em algumas circunstâncias essas taxas e proporções incluem zeros e/ou uns com probabilidade positiva. Nesses casos não é mais possível supor que os dados provêm de uma distribuição estritamente contínua. Dessa forma, há necessidade de modelar esses dados por meio de distribuições mistas para acomodar dados limitados no intervalo $(0,1]$, $[0,1)$ ou $[0,1]$. A distribuição beta inflacionada, proposta por Ospina e Ferrari (2010), é uma distribuição de probabilidade mista, sendo ela contínua e discreta. Para dados observados no intervalo $(0,1]$ ou $[0,1)$ é utilizado uma mistura da distribuição beta para $y \in (0, 1)$ e uma distribuição degenerada que atribui probabilidade não negativa a 0 ou 1. Para dados observados em $[0,1]$ também é utilizado uma mistura da distribuição beta para $y \in (0, 1)$ e a distribuição Bernoulli (BAYES; VALDIVIESO, 2016).

Em algumas situações, as taxas e proporções a serem monitoradas podem ser correlacionadas com uma ou mais variáveis auxiliares, e assim, a utilização de gráficos de controle convencionais pode resultar em distorções na taxa de falsos alarmes, reduzindo assim a eficiência do processo. Essa é a situação em que as causas normais, que atuam de forma aleatória, são confundidas com causas especiais, que não seguem um padrão aleatório,

chegando assim, a uma conclusão errônea quanto a estabilidade do processo. Para isso, uma ferramenta de controle eficaz proposta por Mandel (1969) é o gráfico de controle baseado em regressão. Esse método é utilizado quando a variável resposta é relacionada com uma ou mais variáveis auxiliares.

Uma alternativa proposta por Lima-Filho et al. (2020) para modelar dados no intervalo $(0,1]$ ou $[0,1)$ que são correlacionadas com variáveis auxiliares é o gráfico de controle de regressão beta inflacionado (IBRCC). Esse gráfico, do tipo Shewhart, é muito utilizado principalmente quando o objetivo é detectar grandes variações na média do processo. Porém, são insensíveis para pequenas mudanças, perdendo eficiência na identificação desse tipo de variação. Page (1961) propôs a utilização do gráfico de controle de somas cumuladas, mais conhecido como CUSUM, que é mais sensível na detecção de pequenas variações no processo, uma vez que, guarda informações acumuladas da sequência de pontos.

Há alguns estudos na literatura sobre o uso de regressão nos gráficos de controle CUSUM. Chen e Huang (2014) estudaram o uso do gráfico de controle CUSUM baseado em resíduos para detectar o surto da síndrome respiratória no Taiwan. Tan et al. (2021) utilizaram o gráfico de controle de soma cumulativa baseado na regressão de Poisson inflacionada em zero para monitorar dados da gripe. Hwang (2021) propôs o gráfico de controle CUSUM e o gráfico de controle EWMA baseados no desvio residual para monitorar a média do processo com distribuição beta.

Com a limitação do gráfico de controle do tipo Shewhart em detectar pequenas mudanças na média do processo, o estudo propõe um gráfico de controle da soma cumulativa baseado na regressão beta inflacionada para monitorar dados limitados correlacionados com uma ou mais variáveis auxiliares. O desempenho do gráfico de controle proposto é comparado com a performance do gráfico IBRCC, em termos de ARL, MRL e SDRL, para verificar a sensibilidade de cada gráfico de controle em detectar mudanças no processo.

1.1 Organização do Trabalho

Após essa seção introdutória, esse trabalho está organizado da seguinte forma. A seção 2 apresenta o modelo de regressão beta inflacionado, assim como a estrutura do resíduo Pearson utilizado no estudo. Os gráficos de controle presentes no trabalho, sendo eles o gráfico proposto e o gráfico de controle de regressão beta inflacionado, são discutidos na seção 3. As descrições dos procedimentos de simulação, seus resultados, a análise da sensibilidade e as medidas de desempenho ARL, MRL e SDRL para avaliar a performance dos gráficos de controle são apresentados na seção 4, enquanto uma aplicação a dados reais

é apresentada na seção 5 e a conclusão na seção 6.

CAPÍTULO 2

REGRESSÃO BETA INFLACIONADA

Nesse estudo foram considerados dados limitados como taxas e proporções. Na prática, muitas vezes esses dados podem incluir zeros e/ou uns, não mais apresentando dados no intervalo (0,1). Diante disso, utilizar uma distribuição estritamente contínua como, por exemplo, a beta, não é mais adequado.

Bayes e Valdivieso (2016) reparametrizaram a modelagem estatística proposta por Ospina e Ferrari (2010) para dados distribuídos de forma contínua no intervalo (0, 1), mas que incluem observações em um ou em ambos os extremos. Visto que nesse estudo foram considerados dados limitados definidos nos intervalos [0,1) ou (0,1], a distribuição beta inflacionada foi utilizada.

A função densidade de probabilidade (fdp) da distribuição beta inflacionada é dada por:

$$f(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} \alpha_0(1 - \gamma), & \text{se } y = 0, \\ c\mathcal{B}(y; \mu, \phi), & \text{se } y \in (0, 1), \\ \alpha_1\gamma, & \text{se } y = 1 \end{cases} \quad (2.1)$$

em que $0 < \alpha_0 < 1$, $0 < \alpha_1 < 1$, $0 < \gamma < 1$ e $\phi > 0$ são os parâmetros da distribuição, sendo ϕ o parâmetro de precisão, γ a média da variável aleatória, α_0 é o parâmetro relacionado a $P(y = 0)$ tal que $P(y = 0) = \alpha_0(1 - \gamma)$ e α_1 é o parâmetro relacionado a $P(y = 1)$ tal que $P(y = 1) = \alpha_1\gamma$. Já $c = 1 - \alpha_0(1 - \gamma) - \alpha_1\gamma$ e $\mu = \gamma(1 - \alpha_1)/c$. A função de densidade beta $\mathcal{B}(y; \mu, \phi)$ é definida por:

$$\mathcal{B}(y; \mu, \phi) = f(y; \alpha_0, \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1 \quad (2.2)$$

A partir disso, tem-se que a média e a variância dessa distribuição são dadas por:

$$E(y) = \gamma \quad (2.3)$$

$$\text{Var}(y) = \frac{(1 + \alpha_1\phi)}{1 + \phi} \gamma + \left(\frac{(1 - \alpha_1)^2\phi}{(1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)(1 + \phi)} - 1 \right) \gamma^2 \quad (2.4)$$

Note que, se $\alpha_0 = 0$ e $\alpha_1 > 0$, a distribuição é conhecida como beta inflacionada em 1 e representada por $Y \sim \text{BEOI}(\alpha, \gamma, \phi)$. Em contrapartida, se $\alpha_1 = 0$ e $\alpha_0 > 0$, então é chamada de beta inflacionada em 0, simbolizada por $y \sim \text{BEZI}(\alpha, \gamma, \phi)$.

2.1 Modelo de Regressão

O modelo de regressão beta inflacionado com dispersão variável é dado pela seguinte estrutura:

$$g_1(\alpha_{0i}) = \tilde{x}_i^\top \omega \quad (2.5)$$

$$g_2(\alpha_{1i}) = \check{x}_i^\top \kappa \quad (2.6)$$

$$g_3(\gamma_i) = x_i^\top \beta \quad (2.7)$$

$$g_4(\phi_i) = \ddot{x}_i^\top \zeta \quad (2.8)$$

em que $i = 1, \dots, n$, $\omega = [\omega_0, \omega_1, \dots, \omega_{k_1}]^\top$, $\kappa = [\kappa_0, \kappa_1, \dots, \kappa_{k_2}]^\top$, $\beta = [\beta_0, \beta_1, \dots, \beta_{k_3}]^\top$ e $\zeta = [\zeta_0, \zeta_1, \dots, \zeta_{k_4}]^\top$ são vetores de parâmetros de regressão desconhecidos. No estudo eles são estimados por meio do método de máxima verossimilhança apresentado no capítulo

seguinte. Já $\tilde{x}_i = [1, \tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ik_1}]^\top$, $\check{x}_i = [1, \check{x}_{i1}, \check{x}_{i2}, \dots, \check{x}_{ik_2}]^\top$, $x_i = [1, x_{i1}, x_{i2}, \dots, x_{ik_3}]^\top$ e $\ddot{x}_i = [1, \ddot{x}_{i1}, \ddot{x}_{i2}, \dots, \ddot{x}_{ik_4}]^\top$ são vetores colunas de variáveis independentes e $g_1 : (0, 1) \rightarrow \mathbb{R}$, $g_2 : (0, 1) \rightarrow \mathbb{R}$, $g_3 : (0, 1) \rightarrow \mathbb{R}$ e $g_4 : (0, \infty) \rightarrow \mathbb{R}$ são funções de ligação estritamente monótonas e duas vezes diferenciáveis, em que g_1 é a estrutura de regressão para α_0 que é o parâmetro relacionado a proporção de zeros, g_2 é a estrutura de regressão para α_1 que é o parâmetro relacionado a proporção de uns, g_3 é a estrutura de regressão para a média e g_4 para o parâmetro de precisão. Enquanto que para os submodelos de α_0 , α_1 e γ podemos usar diferentes funções de ligações como logit, log-log ou probit, para a estrutura do ϕ é possível usar log ou raiz quadrada (LIMA-FILHO et al., 2020). Para seleção das variáveis foi considerado um nível de significância de 5% conforme apresentado no capítulo das aplicações.

2.2 Resíduo

Os resíduos têm um papel importante na verificação da adequação do modelo, investigando características que comprometem a validade do modelo e avaliando a qualidade do seu ajuste. Além disso, a análise de resíduo é uma ótima forma de identificar observações discrepantes, chamadas de outliers, e observações influentes. De forma genérica, r_i representa a diferença entre o valor observado (y_i) e o respectivo valor estimado pelo modelo (\hat{y}_i).

$$r_i = y_i - \hat{y}_i \quad (2.9)$$

Há diferentes tipos de resíduos possíveis a serem utilizados para o modelo de regressão beta inflacionado, como por exemplo o resíduo quantil aleatorizado, resíduo padronizado e o resíduo Pearson. No estudo foram analisados alguns resíduos, porém o que obteve maior desempenho foi o resíduo de Pearson. Dessa forma, dado a sua simplicidade, nesse trabalho foi utilizado o resíduo de Pearson, com característica de distribuição assimétrica para modelos não-normais. De acordo com Lai et al. (2021), o resíduo de Pearson é a medida mais comum de adequação e pode ser expressa por:

$$r_i^{Pea} = \frac{y_i - \hat{\gamma}_i}{\widehat{Var}(y_i)} \quad (2.10)$$

em que $\widehat{Var}(y_i)$ é a variância de y dada em 2.4.

CAPÍTULO 3

GRÁFICOS DE CONTROLE

O conceito elaborado pelo Dr. Walter Andrew Shewhart em 1924 conhecido como gráfico de controle é uma das principais ferramentas do CEP para o monitoramento do processo. Esse tipo de instrumento é utilizado para identificar se o processo está sob controle ou fora de controle. Além disso, é um método para a obtenção da estabilidade do processo, melhorias por meio da redução da variabilidade e otimização dos processos. É uma forma de apresentar visualmente os dados do processo, além de ser uma técnica muito útil para o monitoramento do processo, revelando quando necessita de ações corretivas por apresentar alterações.

A notoriedade do gráfico de controle é grande pois é uma técnica considerada simples e eficiente. Segundo Montgomery (2009), há pelo menos cinco razões pela sua popularidade:

- É uma técnica comprovada para a melhoria da produtividade;
- É eficiente na prevenção de defeitos;
- Evita o ajuste desnecessário do processo;
- Fornece informações de diagnóstico do processo;
- Fornece informações sobre a capacidade do processo.

Os gráficos de controle são muito poderosos para monitorar as variações dos produtos. Sua visualização é dada por meio de um conjunto de amostras, sendo interpretadas através de três linhas horizontais. De acordo com a Figura 3.1, a linha central (LC), que representa

o valor médio da característica do processo. Há também o limite superior de controle (LSC) e o limite inferior de controle (LIC).

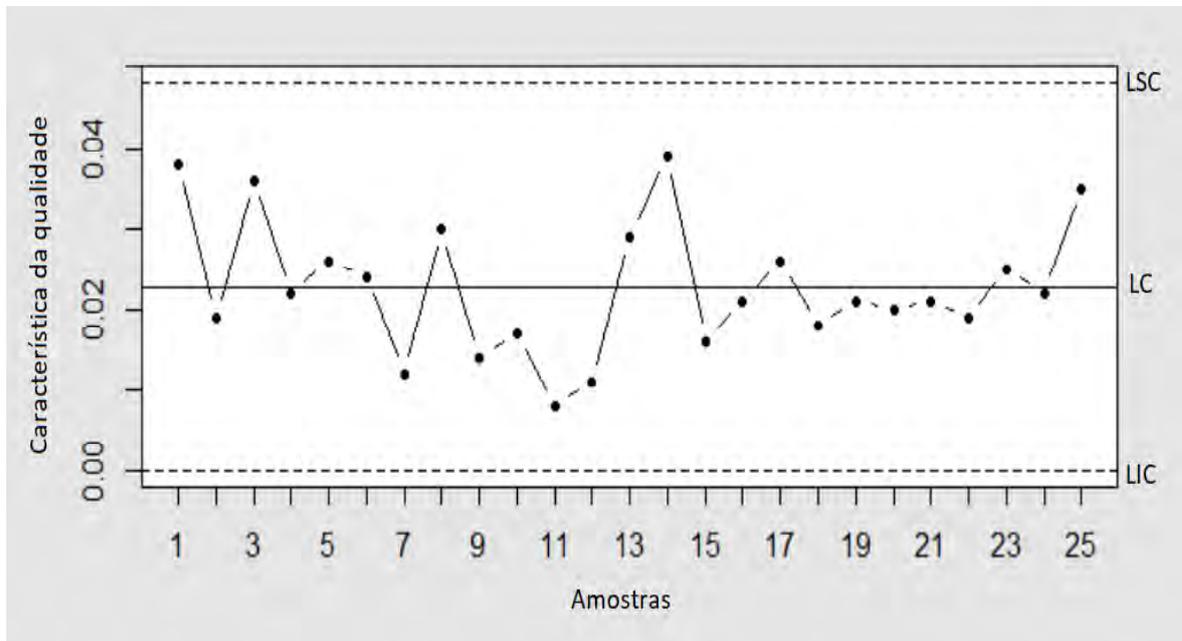


Figura 3.1: Gráfico de controle.

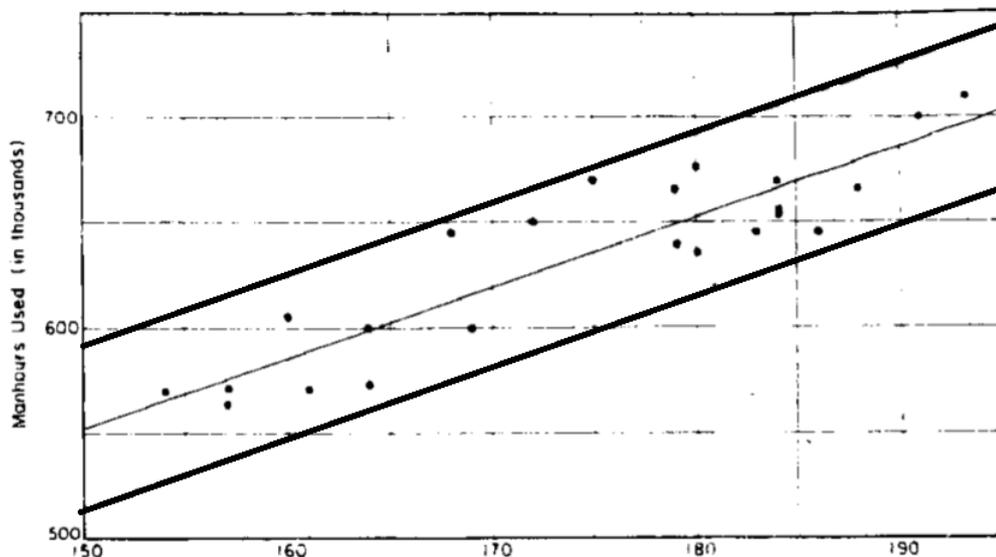
Quando o processo está sob controle, os pontos devem estar entre esses dois limites, superior e inferior, ou seja, dentro dos parâmetros desejados, comprovando que nenhuma ação é necessária. Conforme é possível observar no exemplo da Figura 3.1, todas as amostras se apresentam dentro dos limites delimitados, evidenciando assim que o processo está sob controle. Se for observado algum ponto fora desses limites, é um sinal de anormalidade e que o processo está fora de controle, precisando de ações corretivas para que o motivo desse desvio seja identificado e sanado.

Quando os dados estudados não são correlacionados com uma ou mais variáveis auxiliares, o uso do gráfico tradicional pode ser uma boa solução para monitorar o processo. Porém, em muitas situações práticas a característica da qualidade está relacionada com uma ou mais variáveis auxiliares. Nesses casos, a utilização de gráficos convencionais pode ocasionar em falsos alarmes, podendo assim concluir de forma errônea que o processo está fora de controle quando na verdade são causas naturais. A seguir são apresentados dois gráficos de controle estudados para monitorar processos limitados inflacionados em zero ou um que fazem uso de variáveis auxiliares.

3.1 Gráfico de controle para regressão

A utilização do gráfico de controle baseado em um modelo de regressão foi proposta por Mandel (1969). Esse método é utilizado quando temos uma variável auxiliar correlacionada com a característica da qualidade, controlando a variação média da variável de interesse ao invés de controlar uma média constante como é feito nos gráficos de controle tradicionais. Conforme apresentado na Figura 3.2, as linhas do gráfico de controle de regressão linear simples são inclinadas em relação ao eixo horizontal, sendo paralelas à linha central. O uso desse gráfico de controle só é possível em processos com uma ou mais variáveis auxiliares.

Figura 3.2: Gráfico de controle de regressão.



Fonte: Mandel (1969)

3.1.1 Gráfico de controle de regressão beta inflacionado

Algumas alternativas da utilização desse gráfico de controle têm sido propostas na literatura. Dentre elas, Lima-Filho et al. (2020) propuseram o uso do modelo de regressão beta inflacionado para monitorar características da qualidade relacionadas com variáveis auxiliares e limitadas aos intervalos $(0,1]$ ou $[0,1)$. Os limites de controle, considerando esse modelo, são dados por:

$$LIC_i = F^{-1}(\alpha/2; \alpha_{0i}, \alpha_{1i}, \gamma_i, \phi_i) \quad (3.1)$$

$$LC_i = \gamma_i \quad (3.2)$$

$$LSC_i = F^{-1}(1 - \alpha/2; \alpha_{0i}, \alpha_{1i}, \gamma_i, \phi_i) \quad (3.3)$$

em que F^{-1} é a função quantílica da distribuição beta inflacionada. Os parâmetros α_{0i} , α_{1i} , γ_i e ϕ_i são funções de ω , κ , β e ζ , respectivamente, em que $\alpha_{0i} = \mathbf{g}_1^{-1}(\tilde{x}_i^T \omega)$, $\alpha_{1i} = \mathbf{g}_2^{-1}(\tilde{x}_i^T \kappa)$, $\gamma_i = \mathbf{g}_3^{-1}(x_i^T \beta)$ e $\phi_i = \mathbf{g}_4^{-1}(\tilde{x}_i^T \zeta)$.

Na prática, os parâmetros do modelo, α_{0i} , α_{1i} , γ_i e ϕ_i , são desconhecidos e algum método de estimação é necessário para estimar os limites de controle. Para isso, pelas boas propriedades, os estimadores de máxima verossimilhança (MLE) são muito utilizados. A função de log-verossimilhança da distribuição beta inflacionada é dada por:

$$l(\theta) = \sum_{i=1}^n l_i(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi_i) \quad (3.4)$$

$$l(\theta) = \sum_{i=1}^n \log[\alpha_{0i}(1 - \gamma_i)] \mathbf{1}_0(y_i) + \log(\alpha_{1i}\gamma_i) \mathbf{1}_1(y_i) + [\log(c_i) + \log(\mathcal{B}(y_i; \mu_i, \phi_i))][(1 - \mathbf{1}_0(y_i))(1 - \mathbf{1}_1(y_i))] \quad (3.5)$$

em que $\theta = (\omega^T, \kappa^T, \beta^T, \zeta^T)^T$, $\mathbf{1}_0$ é uma função indicadora igual a 1 se $y = 0$ e igual a 0 se $y \in (0, 1]$, $\mathbf{1}_1$ também é uma função indicadora igual a 1 se $y = 1$ e igual a 0 se $y \in [0, 1)$ e

$$\log(\mathcal{B}(y_i; \mu_i, \phi_i)) = \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) y_i^* + (\phi_i - 2) y_i^\dagger \quad (3.6)$$

em que α_{0i} , α_{1i} , γ_i e ϕ_i são dados pela estrutura de regressão em 2.5 a 2.8, respectivamente. Além disso,

$$y_i^* = \begin{cases} \log\left(\frac{y_i}{1-y_i}\right), & y_i \in (0, 1), \\ 0, & y_i = 0 \text{ ou } y_i = 1 \end{cases} \quad (3.7)$$

$$y_i^\dagger = \begin{cases} \log(1 - y_i), & y_i \in (0, 1), \\ 0, & y_i = 0 \text{ ou } y_i = 1 \end{cases} \quad (3.8)$$

Derivando a função log-verossimilhança apresentada em 3.4 com relação a cada elemento do vetor θ , obtemos o vetor Escore dado por $U(\theta) = (U_\omega(\theta)^\top, U_\kappa(\theta)^\top, U_\beta(\theta)^\top, U_\zeta(\gamma)^\top)^\top$. O MLE de θ é obtido resolvendo o sistema não linear $(U_\omega(\theta)^\top, U_\kappa(\theta)^\top, U_\beta(\theta)^\top, U_\zeta(\gamma)^\top)^\top = 0$, onde 0 é um vetor nulo de dimensões $(k_1 + k_2 + k_3 + k_4)$. Como os MLE's não possuem forma fechada, são obtidos numericamente conduzidos através de um algoritmo de Newton. Neste trabalho, usamos o Broyden-Fletcher-Goldfarb-Shanno (BFGS) para implementação computacional. O vetor escore $U(\theta)$ e a matriz de Informação de Fisher podem ser encontrados em Lima-Filho et al. (2020).

Considerando o modelo proposto por Lima-Filho et al. (2020), seus limites de controle estimados ($\widehat{\text{LIC}}$ e $\widehat{\text{LSC}}$), são dados por:

$$\widehat{\text{LIC}}_i = F^{-1}(\alpha/2; \widehat{\alpha}_{0i}, \widehat{\alpha}_{1i}, \widehat{\gamma}_i, \widehat{\phi}_i) \quad (3.9)$$

$$\widehat{\text{LC}}_i = \widehat{\gamma}_i \quad (3.10)$$

$$\widehat{\text{LSC}}_i = F^{-1}(1 - \alpha/2; \widehat{\alpha}_{0i}, \widehat{\alpha}_{1i}, \widehat{\gamma}_i, \widehat{\phi}_i) \quad (3.11)$$

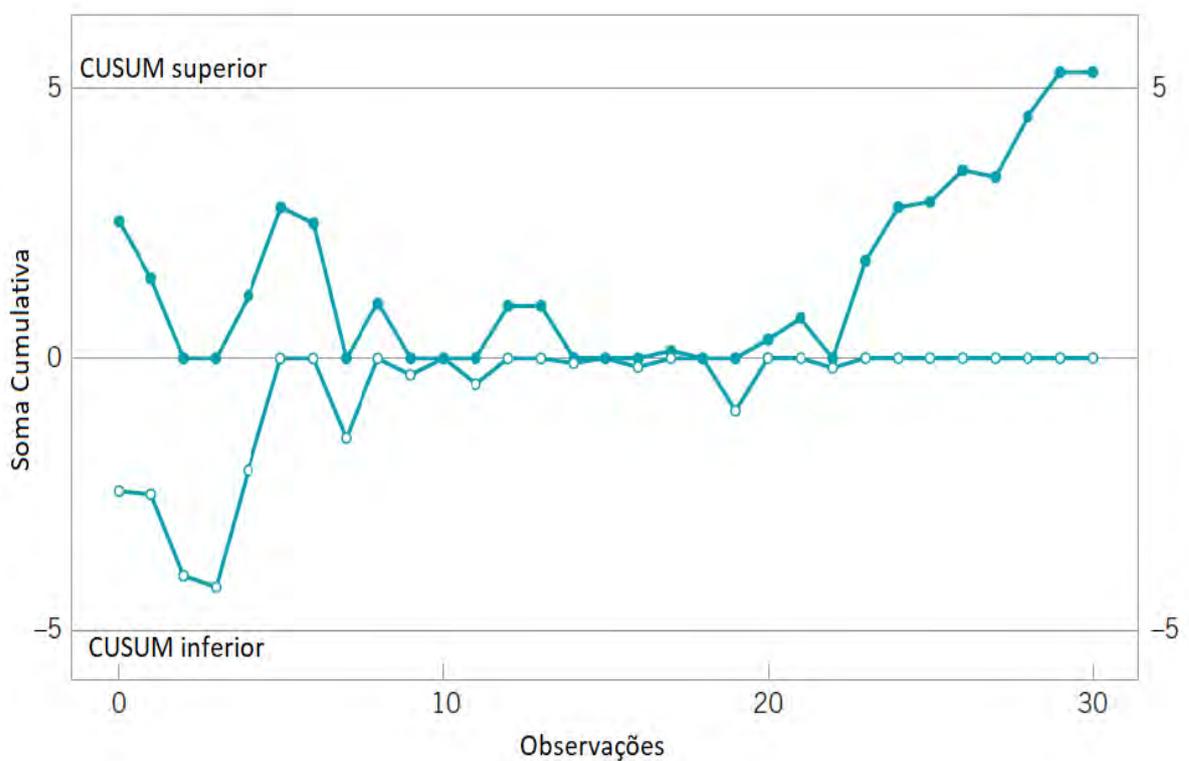
em que $\widehat{\alpha}_{0i} = g_1^{-1}(\tilde{x}_i^T \widehat{\omega})$, $\widehat{\alpha}_{1i} = g_2^{-1}(\tilde{x}_i^T \widehat{\kappa})$, $\widehat{\gamma}_i = g_3^{-1}(x_i^T \widehat{\beta})$ e $\widehat{\phi}_i = g_4^{-1}(\tilde{x}_i^T \widehat{\zeta})$.

Gráficos de controle do tipo Shewhart são muito úteis para detectar grandes variações na média do processo, sendo popular por conta da sua simplicidade e facilidade de interpretação. Esse tipo de gráfico também é conhecido pelo fato de não ter memória por considerar apenas a observação em um determinado momento, ou seja, ignora qualquer informação dada pela sequência inteira de pontos. Isso faz com que o gráfico de controle perca eficiência por não levar em conta observações anteriores. Tais características tornam esse tipo de gráfico insensível a pequenas mudanças no processo. Para manter um controle mais seguro e rigoroso do processo, é fundamental a utilização de gráficos de controle adequado que possam identificar corretamente pequenas mudanças.

3.2 CUSUM baseado no resíduo de Pearson do modelo de regressão beta inflacionado

Em decorrência da dificuldade encontrada nos gráficos do tipo Shewhart em detectar pequenos desvios, houve necessidade de criação de um gráfico de controle que fosse ideal para essas situações. Dessa forma, Page (1961) desenvolveu um gráfico capaz de identificar esse tipo de variação uma vez que acumula as informações das amostras. Esse gráfico ficou conhecido como gráfico de controle de soma cumulada (CUSUM).

Figura 3.3: Gráfico da soma cumulativa.



Fonte: Montgomery (2009)

A Figura 3.3, retirada do livro *Introduction to Statistical Quality Control*, apresenta o comportamento do gráfico CUSUM. Essas estatísticas apresentadas na Figura são conhecidas como CUSUM superior e inferior. Quando um ponto for plotado abaixo do CUSUM inferior ou acima do CUSUM superior, aquela observação será considerada fora de controle e o motivo dessa variabilidade deve ser investigado e sanado. No caso da Figura 3.3 apresentada, todos os pontos estão plotado dentro os limites definidos. Dessa forma, o processo está sob controle.

De acordo com Woodall (1986) esse tipo de gráfico de controle pode detectar pequenas

mudanças na característica da qualidade e ainda assim manter um controle seguro do processo. Além disso, esse gráfico proporciona um controle mais rigoroso do processo do que o gráfico de controle do tipo Shewhart (LUCAS, 1985).

O gráfico CUSUM proposto nesse estudo foi baseado no resíduo de Pearson do modelo de regressão beta inflacionado apresentado na equação 2.10. Dessa forma, a estatística CUSUM-IBRCC é dada por:

$$C_i^+ = \text{máx}[0, r_i^{Pea} - (\mu_0 + K) + C_{i-1}^+] \quad (3.12)$$

$$C_i^- = \text{máx}[0, (\mu_0 - K) - r_i^{Pea} + C_{i-1}^-] \quad (3.13)$$

em que a estatística C_i^+ é a soma acumulada dos desvios positivos, e C_i^- a soma acumulada dos desvios negativos. Ou seja, o CUSUM incorpora diretamente toda informação na sequência dos valores da amostra, plotando as somas cumulativas dos desvios dos valores da amostra de um valor-alvo. Os valores iniciais, C_0^+ e C_0^- , são considerados iguais a 0. Temos ainda que r_i^{Pea} é a i -ésima observação do resíduo de Pearson, μ_0 é a média do resíduo de Pearson e K é um valor de referência sendo aproximadamente $K = \sigma * k$, em que σ é o desvio padrão e o $k = \frac{1}{2} * \lambda$, λ sendo o tamanho da mudança que se deseja detectar em unidades de desvios padrão. Por padrão, nesse trabalho consideramos $\lambda = 1$.

O intervalo de decisão do gráfico de controle CUSUM é expresso como $H = \sigma * h$, em que h é um valor estimado de forma a obter um valor de α desejado. A seleção desse parâmetro deve ser feita cuidadosamente, visto que impacta diretamente na performance e eficiência do CUSUM. Nesse trabalho o valor de h será estimado por meio de uma simulação de Monte Carlo.

4.1 Estudo de Simulação

Foram considerados 6 diferentes cenários, os 3 primeiros cenários foram construídos considerando o modelo de regressão beta inflacionado em zero, sendo eles os cenários 1-3. Os outros 3 cenários foram definidos considerando o modelo de regressão beta inflacionado em um, sendo eles os cenários 4-6. Na Tabela 4.1 observam-se os valores dos parâmetros para os 6 cenários. A primeira coluna é referente aos diferentes cenários avaliados no estudo. Já a segunda e terceira colunas correspondem aos valores para a estrutura do α_0 , enquanto as colunas 4 e 5 são relativas aos valores para estrutura do α_1 . As colunas 6 e 7 são para a estrutura da média (γ), e as colunas 8 e 9 são referentes a estrutura do parâmetro de precisão (ϕ).

Tabela 4.1: Cenários considerados no estudo de simulação

Cenário	ω_0	ω_1	κ_0	κ_1	β_0	β_1	ζ_0	ζ_1
1	-3,0	1,0	0,0	0,0	-2,5	1,5	2,0	-0,5
2	-2,3	0,9	0,0	0,0	-2,3	1,9	2,0	0,7
3	-0,9	-1,0	0,0	0,0	-1,7	-2,0	4,5	-0,3
4	0,0	0,0	-2,8	0,7	3,0	-1,0	3,5	-1,0
5	0,0	0,0	-2,5	1,0	4,0	-1,0	4,5	-0,5
6	0,0	0,0	-3,5	0,5	1,5	-0,5	2,7	-0,7

Foi usada a seguinte estrutura do modelo de regressão beta inflacionado para a geração dos dados:

$$\text{logit}(\alpha_{0i}) = \omega_0 + \omega_1 \tilde{x}_i \quad (4.1)$$

$$\text{logit}(\alpha_{1i}) = \kappa_0 + \kappa_1 \tilde{x}_i \quad (4.2)$$

$$\text{logit}(\gamma_i) = \beta_0 + \beta_1 x_i \quad (4.3)$$

$$\log(\phi_i) = \zeta_0 + \zeta_1 \ddot{x}_i \quad (4.4)$$

em que $i = 1, \dots, n$. Os valores de \tilde{x}_i , \check{x}_i e \ddot{x}_i foram obtidos por meio da distribuição Bernoulli com parâmetro $p = 0,3$. Já x_i foi gerado com base na distribuição uniforme no intervalo $(0,1)$.

A Tabela 4.2 apresenta a característica dos 6 cenários considerados no estudo. A primeira coluna é referente aos cenários, enquanto as demais colunas, 2-5, correspondem à média, precisão, proporção de zeros e proporção de uns, respectivamente. Podemos observar que os cenários 3 e 5 apresentam maior precisão, bem como apresentam maior proporção de zeros e uns, respectivamente. No estudo foram utilizados parâmetros com diferentes característica com o objetivo de representar situações reais.

Tabela 4.2: Característica dos cenários

Cenários	γ	ϕ	proporção de zeros	proporção de uns
1	[0,076; 0,269]	[4,480; 7,39]	[4, 70%; 11, 90%]	-
2	[0,091; 0,401]	[7,390; 14,88]	[9, 10%; 19, 80%]	-
3	[0,024; 0,154]	[66,69; 90,02]	[13, 01%; 28, 91%]	-
4	[0,880; 0,950]	[12,18; 33,12]	-	[5, 70%; 10, 90%]
5	[0,950; 0,980]	[54,60; 90,02]	-	[7, 50%; 18, 20%]
6	[0,730; 0,820]	[7,380; 14,88]	-	[2, 90%; 4, 70%]

O estudo de simulação de Monte Carlo foi utilizado para comparar e avaliar o gráfico proposto (CUSUM-IBRCC usando o resíduo de Pearson) com o IBRCC. De acordo com Schaffer e Kim (2007) 5.000 réplicas de Monte Carlo é suficiente para garantir resultados com boa acurácia. Dito isso, foram geradas 5.000 réplicas considerando $\alpha = 0,01$ e tamanhos de amostra $n = 100, 200$ e 500 na Fase I. Todas as simulações foram realizadas por meio da linguagem de programação R.

Foram considerados diferentes valores de δ , que simboliza a mudança induzida no

processo. Para os cenários inflacionados em zero, o processo encontra-se fora de controle quando $\delta = \{0, 1; 0, 2; 0, 3; 0, 4; 0, 5\}$. Já nos cenários inflacionados em um, o processo encontra-se fora de controle quando o $\delta = \{-0,5; -0,4; -0,3; -0,2; -0,1\}$. O processo encontra-se sob controle para $\delta = 0$ em ambos os casos. Espera-se que, quando considerado um $\alpha = 0,01$, os valores das medidas de desempenho se aproximem dos níveis nominais, ou seja, $ARL_0 = 100$, $MRL_0 = 69$ e $SDRL_0 = 99,5$.

4.1.1 Desempenho dos gráficos de controle

O desempenho do gráfico de controle é normalmente medido em termos do número médio de amostras até que um ponto fora de controle seja detectado. Essa medida de desempenho é conhecida como ARL (*average run length*) (MONTGOMERY, 2009). Outras medidas utilizadas para avaliar o desempenho dos gráficos de controle são a mediana, representada por MRL (*median run length*) e o desvio padrão, representada por SDRL (*standard deviation run length*), da distribuição do comprimento RL (*run length*). São expressas por:

$$ARL = \frac{1}{p} \quad (4.5)$$

$$SDRL = \sqrt{\frac{(1-p)}{p^2}} \quad (4.6)$$

$$MRL = \frac{\ln(0.5)}{\ln(1-p)} \quad (4.7)$$

Essas medidas são avaliadas em duas condições: em controle e fora de controle. Quando o processo está sob controle, essas medidas são dadas em função do erro tipo I e $p = \alpha$. Quando o processo está fora de controle, essas medidas são dadas em função do erro tipo II e $p = 1 - \beta$ (HO; FERNANDES; BOURGUIGNON, 2019). Um gráfico de controle com melhor desempenho é o que mais se aproxima dos níveis nominais em termos de ARL_0 , MRL_0 e $SDRL_0$, em que o subíndice igual a zero indica que o processo está sob controle. Quando os gráficos de controle possuem o mesmo ARL_0 , o gráfico com

melhor desempenho é o que apresenta menor valor de ARL_1 , em que o subíndice igual a um aponta que o processo está fora de controle. Um baixo valor de ARL_1 indica que, quando fora de controle, o processo precisa de poucas observações para sinalizar uma anormalidade (LIMA-FILHO; BAYER, 2021).

4.1.2 Algoritmo

Para avaliar e comparar o desempenho dos gráficos de controle, foram utilizadas as medidas de ARL, MRL e SDRL. A seguir uma breve descrição do procedimento de simulação:

Etapas do algoritmo de simulação:

1. O valor de alpha é fixado em $\alpha = 0,01$;
2. Uma sequência de $n + 10.000$ observações são geradas, sendo $n = 100$, da distribuições uniforme padrão $(0,1)$ e Bernoulli com parâmetro $p = 0,3$;
3. Com os valores gerados no passo 2 juntamente com os valores dos parâmetros definidos na Tabela 4.1 e com uma sequência de $n + 10.000$ observações, calcula-se α_{0i} , γ_i e ϕ_i ;
4. São gerados $n + 10.000$ observações da distribuição beta inflacionada em 0 com α_{0i} , γ_i e ϕ_i obtidos no passo 3;
5. O modelo de regressão beta inflacionado é ajustado com dispersão variável para os dados da Fase I (n), obtendo assim os MLE's $\hat{\omega}$, $\hat{\beta}$ e $\hat{\zeta}$;
6. Usando as covariáveis na Fase II, $\hat{\alpha}_{0i}$, $\hat{\gamma}_i$ e $\hat{\phi}_i$ são estimados;
7. Com os dados da Fase II, de acordo com a equação 2.10, o resíduo de Pearson é calculado;
8. Para $\delta = 0$, o CUSUM-IBRCC é construído para o resíduo obtido no passo 7 para o valor estimado de h ;
9. Cada ponto r_i é plotado com os limites de controle no gráfico obtidos no passo 8. Quando for plotado um ponto fora dos limites delimitados, essa amostra é considerada fora de controle.
10. Com exceção do passo 8, o processo foi repetido para todos os valores de delta;

11. Todo o processo descrito acima foi repetido 5 mil vezes;

Conforme fórmulas apresentadas em 4.5, 4.6 e 4.7, as medidas de desempenho \widehat{ARL} , \widehat{SDRL} e \widehat{MRL} foram calculadas. As etapas descritas acima foram repetidas para $n = 200$ e 500 . Por fim, os passos anteriores foram repetidos para todos os cenários. Para os cenários inflacionados em 1, nos passos de 3 a 6, foram considerados o inflacionamento em 1. Onde se tem α_{0t} foi substituído por α_{1t} e onde se tem ω foi substituído por κ .

Antes de avaliar os gráficos de controle, o valor de h foi estimado utilizando um algoritmo similar ao descrito acima. Uma simulação de Monte Carlo foi utilizada para um grid de valores de h , para cada um dos cenários adotados, com o intuito de verificar o \widehat{ARL} para cada um dos valores de h . A partir daí foi possível estimar o valor de h para o ARL desejado.

4.2 Análise de Sensibilidade

A Tabela A.1 apresenta os resultados das medidas de desempenho \widehat{ARL}_0 , \widehat{MRL}_0 e \widehat{SDRL}_0 para os 6 cenários considerados. Os 3 primeiros cenários são referentes à performance dos gráficos de controle CUSUM-IBRCC e IBRCC quando considerado o modelo de regressão beta inflacionado em 0. Já os 3 últimos cenários, são referentes aos gráficos de controle CUSUM-IBRCC e IBRCC quando considerado o modelo de regressão beta inflacionado em 1.

Avaliando a Tabela A.1, é possível observar que os resultados dos dois gráficos de controle, CUSUM-IBRCC e IBRCC, foram semelhantes quando levado em conta o processo em controle. Os valores das medidas de desempenho se aproximam do valor teórico quando $\delta = 0$ em ambos os casos. Contudo, o CUSUM ainda assim apresentou resultados levemente melhores do que o IBRCC, retratando valores de \widehat{ARL}_0 mais próximos do valor nominal.

Como esperado, dentre os gráficos de controle avaliados, o que apresentou melhores resultados de performance para detecção de pequenas mudanças no processo foi o CUSUM-IBRCC em todos os cenários considerados (Tabela A.1). Ou seja, o gráfico de controle CUSUM-IBRCC apresenta valores muito inferiores de \widehat{ARL}_1 , \widehat{MRL}_1 e \widehat{SDRL}_1 do que o IBRCC. Essa é uma evidência da eficiência do gráfico CUSUM-IBRCC em detectar pequenas alterações no processo.

As Figuras 4.1 e 4.2 apresentam os valores da medida de desempenho ARL para todos

os cenários considerados. Em todos os casos, a linha do gráfico CUSUM-IBRCC (linha tracejada azul) está abaixo da linha do gráfico de controle IBRCC (linha contínua vermelha). Esse fato evidencia que o gráfico proposto é mais sensível para detectar mudanças no processo quando comparado com o IBRCC. Por exemplo, quando considerado o cenário 1, $\delta = 0,1$ e $n = 100$, o IBRCC leva aproximadamente 79 amostras para detectar uma mudança no processo enquanto o CUSUM-IBRCC leva aproximadamente 27 amostras para detectar uma mudança de mesma magnitude.

Quando é observado o cenário com baixa precisão como o cenário 1, com $\delta = 0,2$ e $n = 500$, por exemplo, o CUSUM-IBRCC leva, em média, aproximadamente 9 amostras para detectar a mudança no processo. Quando é levado em conta um cenário com alta precisão, como o cenário 3, com os mesmos valores de δ e n , o CUSUM-IBRCC leva, em média, aproximadamente 4 amostras para indentificar a alteração no processo, evidenciando assim um desempenho superior. Pelos gráficos 4.1 e 4.2 é possível observar que o cenário 3 é o que apresenta a curva mais acentuada, confirmando que, quando o processo está fora de controle, o cenário com a maior precisão é o que identifica mais rápido uma mudança no processo.

Nota-se que a performance de ambos os gráficos de controle não mudam de acordo com o aumento do tamanho de amostra. Assim, pelos resultados apresentados nas Figuras 4.1 e 4.2, há evidências para afirmar que o aumento do tamanho da amostra na Fase I não melhora a performance do gráfico de controle em detectar mudanças no processo para CUSUM-IBRCC e IBRCC.

Por fim, é importante ressaltar que levando em conta o processo fora de controle, o gráfico de controle CUSUM-IBRCC apresentou performance superior na detecção de pequenas alterações no processo. Dessa forma, esse gráfico de controle pode ser considerado a melhor alternativa para detectar essas mudanças. Por motivo de similaridade, as Figuras referentes às performances dos gráficos de controle em termos de MRL e SDRL não foram apresentadas.

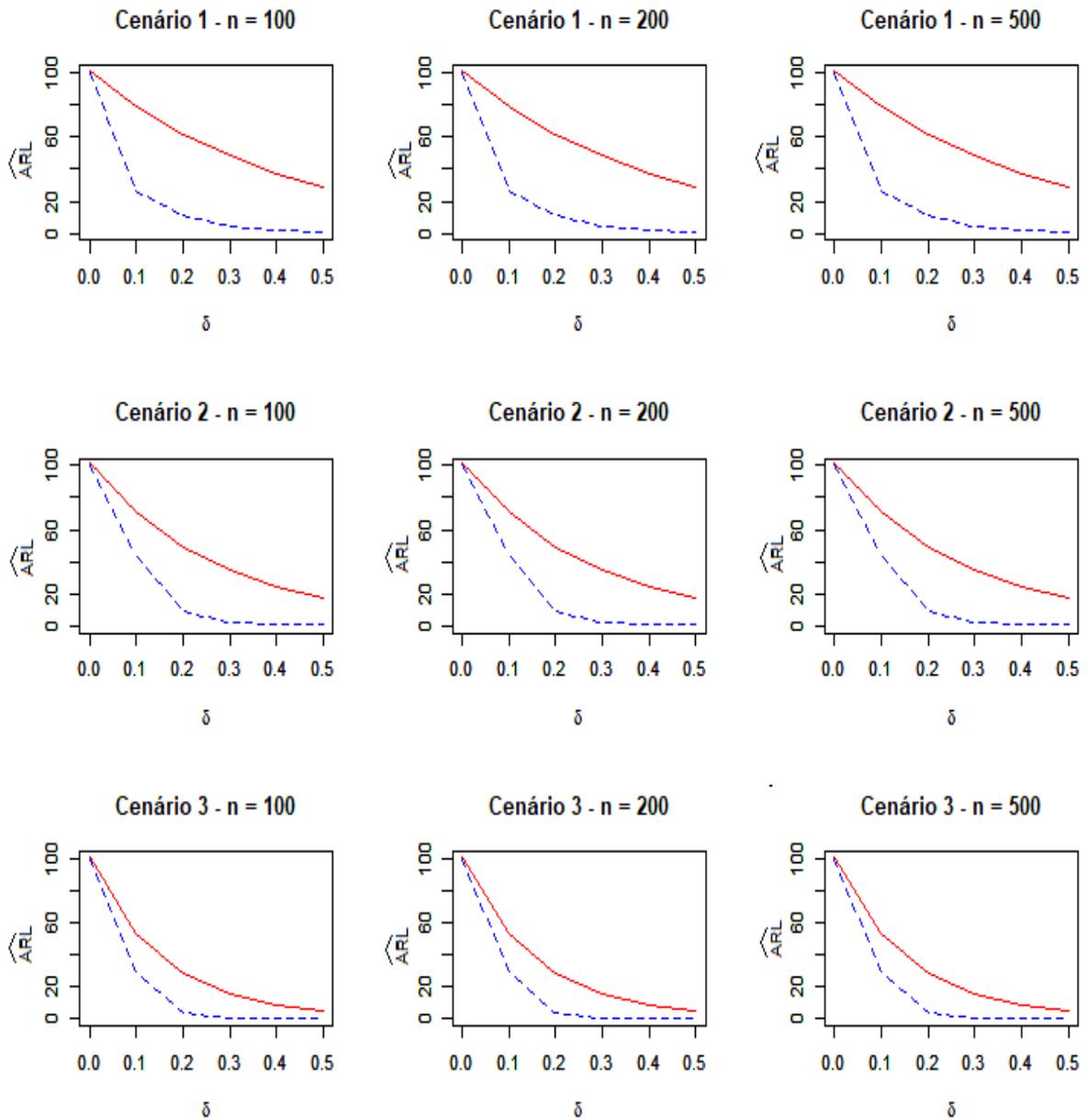


Figura 4.1: Valores de \widehat{ARL} para os cenários 1, 2 e 3 utilizando CUSUM-IBRCC (linha tracejada azul) e IBRCC (linha contínua vermelha)

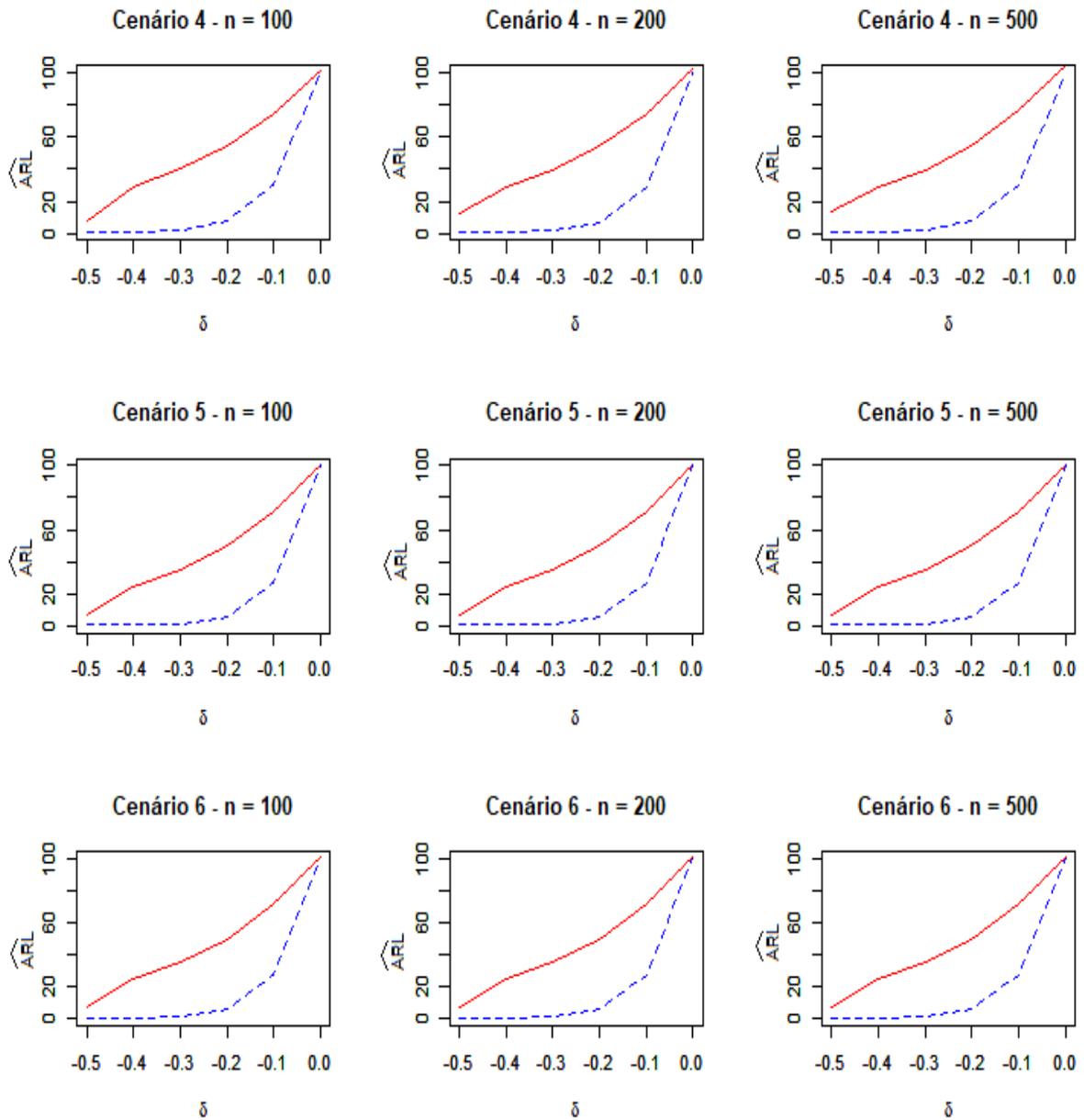


Figura 4.2: Valores de \widehat{ARL} para os cenários 4, 5 e 6 utilizando CUSUM-IBRCC (linha tracejada azul) e IBRCC (linha contínua vermelha)

Nessa seção é apresentada uma aplicação a dados reais. A aplicação é referente à observações meteorológicas de Albury, cidade na Austrália. O banco de dados possui 307 observações entre 2009 e 2019. O banco de dados da aplicação é o weatherAUS, disponível no Software R. Para obtenção dos dados foi utilizado o pacote rattle. A variável monitorada é a umidade relativa às 9 horas da manhã. Dado que essa variável muda ao longo do ano de acordo com as estações, optamos por selecionar o mês de agosto, que corresponde ao inverno. Todas as variáveis utilizadas para ajustar o modelo de regressão estão apresentadas na Tabela 5.1. Além disso, como é possível observar na Figura 5.1 e na Tabela 5.2, a variável de interesse apresenta observações no intervalo $(0,1]$. As 150 primeiras observações são utilizadas para ajustar o modelo de regressão beta inflacionado. As demais observações, 157, são utilizadas na Fase II para o monitoramento do processo. Para essa aplicação é utilizado um $\alpha = 0,01$.

Tabela 5.1: Descrição das variáveis utilizadas no modelo de regressão

Variável	Descrição
Humidity9am	% de umidade relativa às 9am
WindSpeed9am	Média da velocidade do vento (km/h) 10 minutos antes das 9am
Pressure9am	Pressão atmosférica reduzida para o nível médio do mar às 9am
Temp9am	Temperatura em C° às 9am

A Tabela 5.2 apresenta a estatística descritiva das variáveis utilizadas no modelo de regressão. São apresentados os valores mínimo (Min.), primeiro quartil ($Q_{1/4}$), mediana, média, terceiro quartil ($Q_{3/4}$), máximo (Máx.) e o desvio padrão (d.p.) das variáveis. Dessa forma, pela tabela apresentada, a umidade relativa mínima às 9am é de 58,00 e a

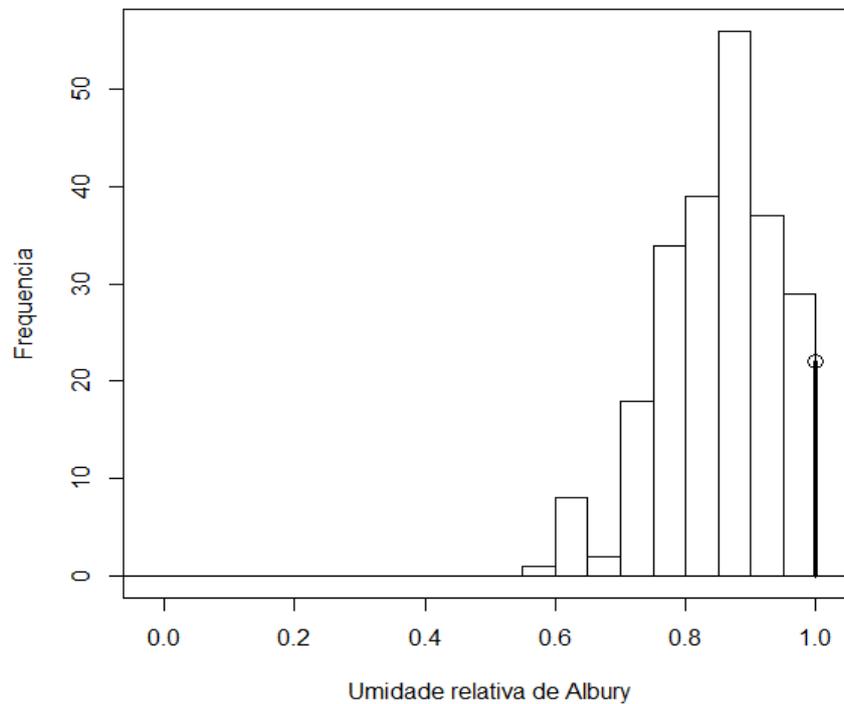


Figura 5.1: Histograma da umidade relativa na cidade Albury - Austrália nos meses de agosto de 2009 a 2019

máxima é de 100,00. Além disso, 75% das observações da variável de interesse apresentam valores maiores ou iguais a 80,00.

Tabela 5.2: Estatística descritiva das variáveis utilizadas no modelo

	Humidity9am	WindSpeed9am	Pressure9am	Temp9am
Min. :	58,00	0,00	1003,00	2,00
Q _{1/4} :	80,00	2,00	1016,00	6,40
Mediana :	87,00	6,00	1021,00	7,80
Média :	86,80	7,70	1021,00	8,11
Q _{3/4} :	94,00	11,00	1026,00	9,70
Máx. :	100,00	35,00	1040,00	15,00
d.p. :	9,42	7,23	7,42	2,56

Conforme expressões apresentadas em 4.1 a 4.4, foi usada a ligação logit para α_1 e γ , enquanto para ϕ foi usada a ligação log. As estimativas de máxima verossimilhança dos parâmetros do modelo são exibidas na Tabela 5.3, considerando 5% de significância.

As Figuras 5.2 e 5.3 apresentam os gráficos CUSUM-IBRCC e IBRCC com os dados da Fase II, considerando $\alpha = 0,01$. Considerando o IBRCC para monitorar a umidade relativa às 9am, o gráfico identificou 7 pontos abaixo do limite inferior, sendo assim, estão

Tabela 5.3: Modelo de regressão ajustado para os dados de umidade relativa

Submodelo para γ			
	Estimado	Erro padrão	p-valor
Intercepto	15,3507	5,7600	0,0077
WindSpeed9am	-0,0163	0,0050	0,0011
Pressure9am	-0,0121	0,0056	0,0312
Temp9am	-0,1168	0,0144	0,0000
Submodelo para ϕ			
	Estimado	Erro padrão	p-valor
Intercepto	2,3519	0,1815	0,0000
WindSpeed9am	0,0315	0,0152	0,0378
Submodelo para α_1			
	Estimado	Erro padrão	p-valor
Intercepto	-34,4699	10,7303	0,0013
Pressure9am	0,0309	0,0105	0,0034

fora de controle. Os pontos fora dos limites delimitados foram as observações 93, 102, 107, 122, 123, 124 e 155. Já o gráfico CUSUM-IBRCC identificou 3 pontos acima do limite superior, sendo eles as observações 49, 50 e 51, e 4 pontos abaixo do limite inferior, sendo eles 124, 125, 126 e 127. Todas as 7 observações citadas foram consideradas fora de controle.

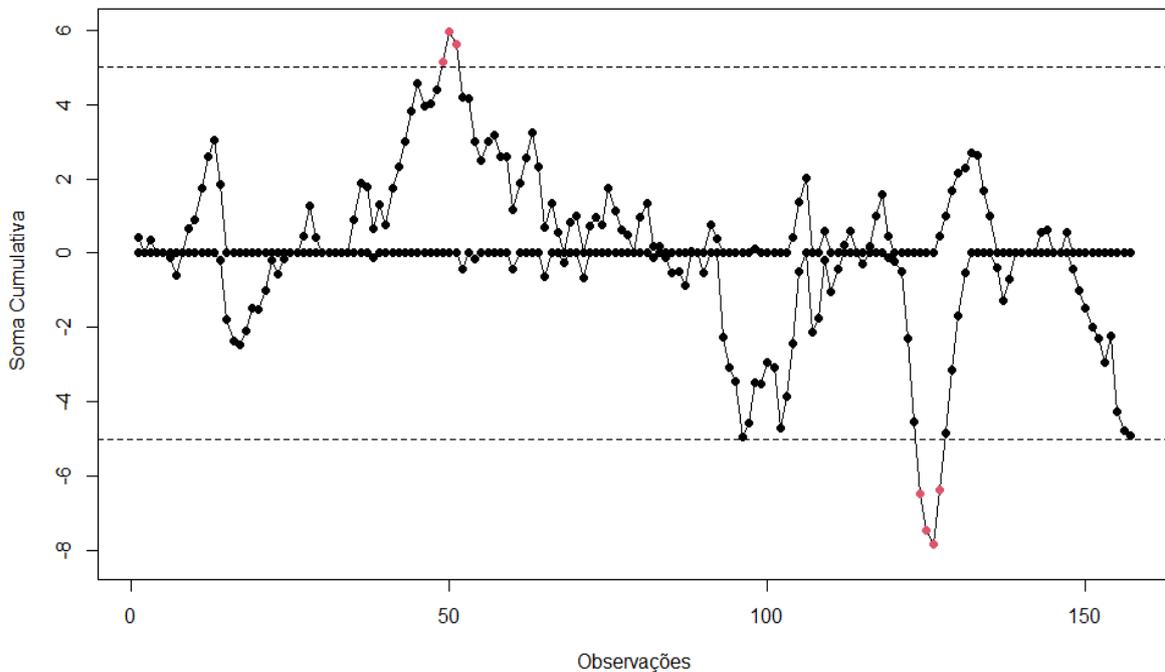


Figura 5.2: Gráfico de controle CUSUM-IBRCC

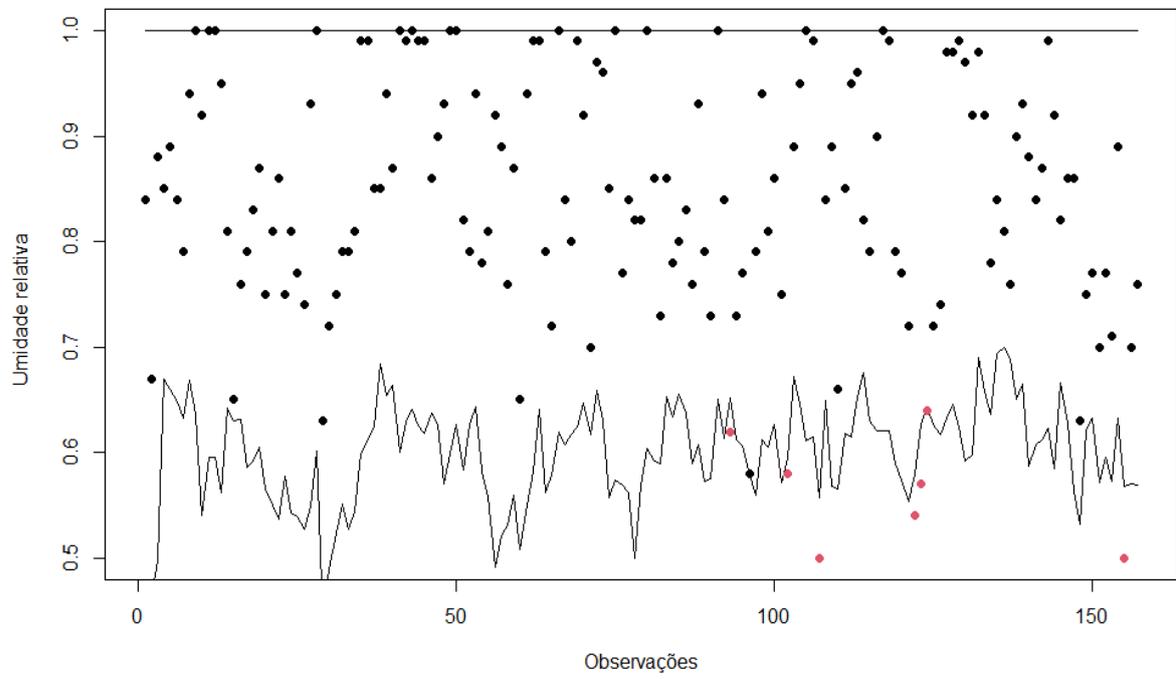


Figura 5.3: Gráfico de controle IBRCC

Nesse estudo foi desenvolvido um novo gráfico de controle para monitorar variáveis limitadas como as taxas e proporções. A proposta do estudo é a utilização do CUSUM considerando o resíduo de Pearson no modelo de regressão beta inflacionado em zero ou um. Este estudo teve como objetivo propor o gráfico de controle CUSUM-IBRCC. Um extenso estudo de simulação Monte Carlo foi realizado para avaliar o desempenho dos gráficos de controle considerando o processo em controle e fora de controle. As medidas de desempenho utilizadas para fazer a comparação dos gráficos foram ARL, MRL e SDRL. Para estimação dos parâmetros na Fase I, foram utilizadas amostras de tamanho $n = 100$, 200 e 500 .

Os resultados obtidos nas simulações sugerem que, quando o processo está sob controle, o gráfico de controle proposto se aproximou mais dos valores nominais das medidas de desempenho do que o IBRCC, ou seja, $ARL_0 = 100$, $MRL_0 = 69$ e $SDRL_0 = 99.5$. Além disso, quando considerado o processo fora de controle, o CUSUM-IBRCC apresentou valores menores de \widehat{ARL}_1 , \widehat{MRL}_1 e \widehat{SDRL}_1 , evidenciando que é mais sensível que o IBRCC, sendo a melhor opção quando o objetivo é detectar pequenas mudanças no processo para manter um controle mais seguro. Além disso, em ambos os gráficos, há evidências de que o aumento do tamanho da amostra na Fase I, não melhora a performance do gráfico de controle em detectar mudanças no processo.

Esse estudo apresenta um novo gráfico de controle para monitorar as variáveis limitadas no intervalo $[0,1)$ ou $(0,1]$ que são correlacionadas com uma ou mais variáveis auxiliares. Diante dos objetivos investigados no estudo, o gráfico proposto mostrou-se uma boa ferramenta no monitoramento desses dados, apresentando bons resultados. Além disso, o

gráfico proposto apresenta desempenho superior ao IBRCC, evidenciando sua eficiência.

REFERÊNCIAS BIBLIOGRÁFICAS

- 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.
- CHAKRABORTI, S.; HUMAN, S.; GRAHAM, M. Phase I statistical process control charts: an overview and some results. *Quality Engineering*, Taylor & Francis, v. 21, n. 1, p. 52–62, 2008.
- CHEN, H.; HUANG, C. The use of a cusum residual chart to monitor respiratory syndromic data. *IIE Transactions*, Taylor & Francis, v. 46, n. 8, p. 790–797, 2014.
- HO, L. L.; FERNANDES, F. H.; BOURGUIGNON, M. Control charts to monitor rates and proportions. *Quality and Reliability Engineering International*, Wiley Online Library, v. 35, n. 1, p. 74–83, 2019.
- HWANG, W.-Y. Deviance residual-based control charts for monitoring the beta-distributed processes. *Quality and Reliability Engineering International*, Wiley Online Library, 2021.
- LAI, X. et al. Residuals based ewma control charts with risk adjustments for zero-inflated poisson models. *Quality and Reliability Engineering International*, Wiley Online Library, 2021.
- LIMA-FILHO, L. M. A.; BAYER, F. Marian o. 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. et al. Process monitoring using inflated beta regression control chart. *Plos one*, Public Library of Science San Francisco, CA USA, v. 15, n. 7, p. e0236756, 2020.
- LUCAS, J. M. Cumulative sum (cusum) control schemes. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 14, n. 11, p. 2689–2704, 1985.
- MANDEL, B. The regression control chart. *Journal of Quality Technology*, Taylor & Francis, v. 1, n. 1, p. 1–9, 1969.

- MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. [S.l.]: Wiley, 2009. v. 6.
- OSPINA, R.; FERRARI, S. L. Inflated beta distributions. *Statistical papers*, Springer, v. 51, n. 1, p. 111–126, 2010.
- PAGE, E. Cumulative sum charts. *Technometrics*, Taylor & Francis, v. 3, n. 1, p. 1–9, 1961.
- SCHAFFER, J. R.; KIM, M.-J. Number of replications required in control chart monte carlo simulation studies. *Communications in Statistics—Simulation and Computation*, Taylor & Francis, v. 36, n. 5, p. 1075–1087, 2007.
- TAN, Y. et al. Risk-adjusted zero-inflated poisson cusum charts for monitoring influenza surveillance data. *BMC Medical Informatics and Decision Making*, BioMed Central, v. 21, n. 2, p. 1–11, 2021.
- WOODALL, W. H. The design of cusum quality control charts. *Journal of Quality Technology*, Taylor & Francis, v. 18, n. 2, p. 99–102, 1986.

APÊNDICE A

RESULTADOS DA SIMULAÇÃO

Tabela A.1: Comparação da performance dos gráficos de controle inflacionados em termos de ARL, MRL e SDRL.

δ	n = 100												n = 200												n = 500															
	IBRCC				CUSUM-IBRCC				IBRCC				CUSUM-IBRCC				IBRCC				CUSUM-IBRCC				IBRCC				CUSUM-IBRCC											
	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL							
0,0	100,98	69,65	100,48	100,04	68,99	99,54	101,98	70,34	101,48	100,01	68,97	99,51	104,98	72,42	104,48	99,99	68,96	99,49	0,1	79,01	54,42	78,50	26,57	18,07	26,06	80,61	61,3	42,14	60,79	82,65	56,94	82,15	20,04	13,54	19,54					
0,1	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49	0,2	61,66	42,39	61,16	11,58	7,68	11,07	63,46	24,01	16,29	63,50	64,85	44,60	64,34	9,20	6,03	8,69					
0,2	49,53	33,98	49,02	9,90	6,51	9,39	47,10	32,30	46,59	15,74	10,56	15,24	53,20	26,03	52,70	14,07	9,40	13,56	0,3	48,09	32,98	47,58	4,84	2,99	4,31	49,92	8,87	5,79	8,35	50,81	34,87	50,31	4,04	2,43	3,50					
0,3	34,83	23,79	34,32	2,30	1,22	1,73	31,88	21,75	31,38	4,26	2,59	3,72	38,05	26,03	37,55	3,50	2,06	2,96	0,4	37,53	25,66	37,02	1,96	0,97	1,37	39,23	3,11	1,79	2,56	39,83	27,26	39,32	1,71	0,79	1,10					
0,4	24,70	16,77	24,20	1,02	0,18	0,15	21,81	14,77	21,30	1,25	0,43	0,55	27,43	18,66	26,92	1,09	0,28	0,32	0,5	29,34	19,99	28,84	1,02	0,18	0,16	30,85	1,12	0,31	0,37	31,22	21,29	30,72	1,06	0,24	0,25					
0,5	17,75	11,96	17,25	1,01	0,16	0,12	15,24	10,21	14,73	1,03	0,19	0,16	20,00	13,51	19,49	1,05	0,23	0,24	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL			
0,0	100,98	69,65	100,48	99,97	68,95	99,47	101,98	70,34	101,48	100,03	68,99	99,53	104,98	72,42	104,48	100,02	68,98	99,52	0,1	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49			
0,1	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49	70,80	48,73	70,30	43,99	30,15	43,49	0,2	49,53	33,98	49,02	9,90	6,51	9,39	47,10	32,30	46,59	15,74	10,56	15,24	53,20	26,03	52,70	14,07	9,40	13,56			
0,2	28,28	19,25	27,78	4,21	2,55	3,67	30,64	20,89	30,14	4,21	2,56	3,68	30,64	20,89	30,14	4,21	2,56	3,68	0,3	34,83	23,79	34,32	2,30	1,22	1,73	31,88	21,75	31,38	4,26	2,59	3,72	38,05	26,03	37,55	3,50	2,06	2,96			
0,3	15,66	10,50	15,15	1,04	0,21	0,19	17,40	11,71	16,89	1,15	0,34	0,41	16,57	11,13	16,06	1,22	0,41	0,52	0,4	9,15	5,99	8,64	1,01	0,16	0,11	10,35	6,82	9,84	1,02	0,18	0,15	9,73	6,39	9,21	1,05	0,23	0,23			
0,4	9,15	5,99	8,64	1,01	0,16	0,11	10,35	6,82	9,84	1,02	0,18	0,15	9,73	6,39	9,21	1,05	0,23	0,23	0,5	5,73	3,61	5,20	1,01	0,16	0,11	6,53	4,17	6,01	1,02	0,18	0,15	6,11	3,88	5,58	1,05	0,23	0,23			
0,5	5,73	3,61	5,20	1,01	0,16	0,11	6,53	4,17	6,01	1,02	0,18	0,15	6,11	3,88	5,58	1,05	0,23	0,23	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
0,0	100,98	69,65	100,48	100,05	69,00	99,55	101,98	70,34	101,48	100,05	69,00	99,55	104,98	72,42	104,48	100,04	68,99	99,54	0,1	52,99	36,38	52,49	29,50	20,10	29,00	55,51	38,13	55,00	19,96	13,48	19,45	55,51	38,13	55,00	19,96	13,48	19,45			
0,1	52,99	36,38	52,49	29,50	20,10	29,00	55,51	38,13	55,00	19,96	13,48	19,45	55,51	38,13	55,00	19,96	13,48	19,45	0,2	28,28	19,25	27,78	4,21	2,55	3,67	30,64	20,89	30,14	4,21	2,56	3,68	30,64	20,89	30,14	4,21	2,56	3,68			
0,2	28,28	19,25	27,78	4,21	2,55	3,67	30,64	20,89	30,14	4,21	2,56	3,68	30,64	20,89	30,14	4,21	2,56	3,68	0,3	15,66	10,50	15,15	1,04	0,21	0,19	17,40	11,71	16,89	1,15	0,34	0,41	16,57	11,13	16,06	1,22	0,41	0,52			
0,3	15,66	10,50	15,15	1,04	0,21	0,19	17,40	11,71	16,89	1,15	0,34	0,41	16,57	11,13	16,06	1,22	0,41	0,52	0,4	9,15	5,99	8,64	1,01	0,16	0,11	10,35	6,82	9,84	1,02	0,18	0,15	9,73	6,39	9,21	1,05	0,23	0,23			
0,4	9,15	5,99	8,64	1,01	0,16	0,11	10,35	6,82	9,84	1,02	0,18	0,15	9,73	6,39	9,21	1,05	0,23	0,23	0,5	5,73	3,61	5,20	1,01	0,16	0,11	6,53	4,17	6,01	1,02	0,18	0,15	6,11	3,88	5,58	1,05	0,23	0,23			
0,5	5,73	3,61	5,20	1,01	0,16	0,11	6,53	4,17	6,01	1,02	0,18	0,15	6,11	3,88	5,58	1,05	0,23	0,23	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
-0,5	8,21	5,33	7,69	1,01	0,15	0,10	12,65	8,42	12,14	1,02	0,18	0,14	13,14	8,76	12,63	1,05	0,23	0,23	-0,4	20,42	20,04	28,91	1,02	0,13	0,13	28,46	19,38	27,96	1,03	0,19	0,16	28,35	19,31	27,85	1,06	0,24	0,25			
-0,4	20,42	20,04	28,91	1,02	0,13	0,13	28,46	19,38	27,96	1,03	0,19	0,16	28,35	19,31	27,85	1,06	0,24	0,25	-0,3	40,29	27,58	39,79	1,81	0,86	1,21	39,32	26,90	38,81	1,58	0,69	0,95	39,48	27,01	38,97	1,94	0,96	1,35			
-0,3	40,29	27,58	39,79	1,81	0,86	1,21	39,32	26,90	38,81	1,58	0,69	0,95	39,48	27,01	38,97	1,94	0,96	1,35	-0,2	55,01	37,79	54,51	7,82	5,07	7,31	54,22	37,24	53,72	7,03	4,52	6,51	54,93	37,73	54,43	8,04	5,22	7,52			
-0,2	55,01	37,79	54,51	7,82	5,07	7,31	54,22	37,24	53,72	7,03	4,52	6,51	54,93	37,73	54,43	8,04	5,22	7,52	-0,1	74,73	51,45	74,23	29,99	20,44	29,49	74,48	51,28	73,98	28,34	19,29	27,83	76,12	52,41	75,62	30,29	20,64	29,78			
-0,1	74,73	51,45	74,23	29,99	20,44	29,49	74,48	51,28	73,98	28,34	19,29	27,83	76,12	52,41	75,62	30,29	20,64	29,78	0,0	100,76	69,49	100,26	100,02	68,98	99,52	101,76	70,19	101,26	99,98	68,96	99,48	104,75	72,26	104,25	100,00	68,97	99,50			
0,0	100,76	69,49	100,26	100,02	68,98	99,52	101,76	70,19	101,26	99,98	68,96	99,48	104,75	72,26	104,25	100,00	68,97	99,50	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
-0,5	7,52	4,86	7,00	1,01	0,15	0,10	11,12	7,36	10,61	1,02	0,18	0,14	8,15	5,30	7,64	1,05	0,23	0,23	-0,4	24,50	16,63	24,00	1,01	0,16	0,12	24,04	16,31	23,53	1,02	0,18	0,16	23,78	16,14	23,28	1,06	0,24	0,24			
-0,4	24,50	16,63	24,00	1,01	0,16	0,12	24,04	16,31	23,53	1,02	0,18	0,16	23,78	16,14	23,28	1,06	0,24	0,24	-0,3	34,99	23,90	34,48	1,41	0,57	0,77	34,55	23,60	34,05	1,39	0,55	0,74	34,52	23,58	34,01	1,59	0,70	0,97			
-0,3	34,99	23,90	34,48	1,41	0,57	0,77	34,55	23,60	34,05	1,39	0,55	0,74	34,52	23,58	34,01	1,59	0,70	0,97	-0,2	49,94	34,27	49,44	6,34	4,04	5,82	49,68	34,08	49,17	6,22	3,96	5,70	50,18	34,43	49,68	6,80	4,36	6,28			
-0,2	49,94	34,27	49,44	6,34	4,04	5,82	49,68	34,08	49,17	6,22	3,96	5,70	50,18	34,43	49,68	6,80	4,36	6,28	-0,1	71,13	48,95	70,62	26,71	18,16	26,20	71,26	49,05	70,76	26,49	18,01	25,99	72,77	50,05	72,21	27,77	18,90	27,26			
0,0	100,76	69,49	100,26	99,99	68,96	99,48	101,76	70,19	101,26	99,99	68,96	99,49	104,75	72,26	104,25	100,01	68,97	99,50	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
-0,5	11,61	7,70	11,10	1,01	0,15	0,10	7,85	5,09	7,33	1,02	0,18	0,14	12,67	8,43	12,16	1,05	0,23	0,23	-0,4	24,53	16,65	24,02	1,01	0,16	0															

APÊNDICE A

IMPLEMENTAÇÃO COMPUTACIONAL - R

```
##### PACOTES #####
```

```
library(gamlss)
library(mvtnorm)
library(betareg)
library(numDeriv)
library(moments)
library(fBasics)
library(nortest)
library(gamlss.dist)
library(LaplacesDemon)
library(qcc)
```

```
##### PROGRAMA - CONTROL CHART INFLATED BETA MODEL #####
```

```
source("imeanbetaregFB.r")
```

```
f1 <- 100 # Phase I
f2 <- 10000 # Phase II
n <- f1 + f2
R <- 5000 # numero de replicas Monte Carlo
ar11 <- 100
a1 <- 1/ar11
```

```
vdelta <- c(seq(-0.5, 0, 0.1))      # numero de observacoes de delta

# DEFININDO O CEN?RIO
omega <- c(0.0, 0.0)                # para estrutura do alpha0
kapa <- c(-2.8, 0.7)                # para estrutura do alpha1
beta <- c(3.0, -1.0)                # mu
csi <- c(3.5, -1.0)                 # phi

testeGL <- matrix(0,n,R)
t0 <- proc.time()

set.seed(1)
X1 <- rbern(n,0.3)
X2 <- X1
X3 <- runif(n)
X4 <- X1
x1 <- cbind(1, X1)
x2 <- cbind(1, X2)
x3 <- cbind(1, X3)
x4 <- cbind(1, X4)
x11 <- as.matrix(X1)
x22 <- as.matrix(X2)
x33 <- as.matrix(cbind(X3), nrow=n, ncol=2, byrow=TRUE)
x44 <- as.matrix(X4)

# ESTATISTICAS
alpha_1 <- rep(0, R)
alpha_2 <- rep(0, R)
alpha_3 <- rep(0, R)
alpha_4 <- rep(0, R)
alpha_5 <- rep(0, R)
alpha_6 <- rep(0, R)
r1_1 <- rep(0, R)
r1_2 <- rep(0, R)
r1_3 <- rep(0, R)
r1_4 <- rep(0, R)
r1_5 <- rep(0, R)
```

```

r1_6 <- rep(0, R)

# VETORES DOS RESULTADOS
mresults <- c()
mresults1 <- c()
set.seed(1)

# SIMULACAO DE MONTE CARLO
for(delta in vdelta)
{

# VERIFICANDO EM QUE PONTO ESTA A SIMULACAO
print("delta")
print(delta)

# EM CONTROLE -----
eta11_in <- x1%%omega
eta21_in <- x2%%kapa
eta31_in <- x3%%beta
eta41_in <- x4%%csi

# PARAMETROS
alpha01_in <- 0
alpha11_in <- exp(eta21_in)/(1+exp(eta21_in))
gama1_in <- exp(eta31_in)/(1+exp(eta31_in))
phi1_in <- exp(eta41_in)
c1_in <- 1 - alpha01_in*(1-gama1_in)-alpha11_in*gama1_in

# REPARAMETRIZANDO
mu1_in <- gama1_in*(1-alpha11_in)/c1_in
delta01_in <- alpha01_in * (1-gama1_in)
delta11_in <- alpha11_in * (gama1_in)
p21_in <- 1 - delta01_in-delta11_in

```

```

# ESTIMANDO Y EM CONTROLE NA BETA INFLACIONADA EM 1
y_in <- rBEINF1(n, mu = mu1_in, sigma = sqrt(1/(phi1_in+1)),
              nu = delta11_in/p21_in)          # em controle
fit <- ibetareg(y_in[1:f1],X2=x22[1:f1],X3=x33[1:f1],X4=x44[1:f1], diag=F)

# EM CONTROLE - FASE II
#omega1 <- as.numeric(fit$omega)
kapa1 <- as.numeric(fit$kapa)
beta1 <- as.numeric(fit$beta)
csi1 <- as.numeric(fit$csi)
#eta1_in <- x1[(f1+1):n,]%**%omega1
eta2_in <- x2[(f1+1):n,]%**%kapa1
eta3_in <- x3[(f1+1):n,]%**%beta1
eta4_in <- x4[(f1+1):n,]%**%csi1

# PARAMETROS
alpha0_in <- 0
alpha1_in <- exp(eta2_in)/(1+exp(eta2_in))
gama_in <- exp(eta3_in)/(1+exp(eta3_in))
phi_in <- exp(eta4_in)
c_in <- 1 - alpha0_in*(1-gama_in)-alpha1_in*gama_in

# REPARAMETRIZANDO
mu_in <- gama_in*(1-alpha1_in)/c_in
delta0_in <- alpha0_in * (1-gama_in)
delta1_in <- alpha1_in * (gama_in)
p2_in <- 1 - delta0_in-delta1_in

# FORA DE CONTROLE -----

#eta1_out <- x1[(f1+1):n,]%**%omega1
eta2_out <- x2[(f1+1):n,]%**%kapa1
eta3_out <- x3[(f1+1):n,]%**%beta1 + delta
eta4_out <- x4[(f1+1):n,]%**%csi1

```

```

# PARAMETROS
alpha0_out <- 0
alpha1_out <- exp(eta2_out)/(1+exp(eta2_out))
gama_out <- exp(eta3_out)/(1+exp(eta3_out))
phi_out <- exp(eta4_out)
c_out <- 1 - alpha0_out*(1-gama_out)-alpha1_out*gama_out

# REPARAMETRIZANDO
mu_out <- gama_out*(1-alpha1_out)/c_out
delta0_out <- alpha0_out * (1-gama_out)
delta1_out <- alpha1_out * (gama_out)
p2_out <- 1 - delta0_out-delta1_out

# GRAFICOS DE CONTROLE DE SHEWHART
if(alpha1_in*(gama_in) >= a1/2){
  LIC1 <- qBEINF1(a1, mu = mu_in, sigma = sqrt(1/(phi_in+1)),
    nu = delta1_in/p2_in)
  LSC1 <- 1
}
if(alpha1_in*(gama_in) < a1/2){
  LIC1 <- qBEINF1(a1/2, mu = mu_in, sigma = sqrt(1/(phi_in+1)),
    nu = delta1_in/p2_in)
  LSC1 <- qBEINF1(1-(a1/2), mu = mu_in, sigma = sqrt(1/(phi_in+1)),
    nu = delta1_in/p2_in)
}

set.seed(1)

# SIMULACAO DE MONTE CARLO
for(i in 1:R)
{

  # ESTIMANDO Y FORA DE CONTROLE NA BETA INFLACIONADA EM 1
  y_out <- rBEINF1(f2, mu = mu_out, sigma = sqrt(1/(phi_out+1)),

```

```

nu = delta1_out/p2_out) # fora de controle

# IBRCC -----
alpha1 <- sum(y_out<LIC1) + sum(y_out>LSC1)
alpha_1[i] <- alpha1/n
rl_1[i] <- diff(c(0,which(y_out<LIC1|y_out>LSC1)))[1]

# CUSUM -----

residPear_in <- fit$resid1

vary <- (((1+alpha1_in*phi_in)/(1+phi_in))*
        gama_in+((((1-alpha1_in)^2)*phi_in)/
        ((1-alpha0_in*(1-gama_in)-alpha1_in*gama_in)*
        (1+phi_in))-1)*(gama_in^2)

residPear_out = (y_out - gama_in)/sqrt(vary)

# METODO 1
# residuo ponderado padronizado 2
limitesPear2 <- cusum(residPear_out,center=mean(residPear_in),
std.dev=sd(residPear_in), decision.interval=3*2.04,
se.shift=1,plot=FALSE)
alphaPear2 <- length(limitesPear2$violations$upper)+
length(limitesPear2$violations$lower)
alpha_2[i] <- alphaPear2/n

limitesPear3 <- cusum(residPear_out,center=mean(residPear_in),
std.dev=sd(residPear_in), decision.interval=3*2.05,
se.shift=1,plot=FALSE)
alphaPear3 <- length(limitesPear3$violations$upper)+
length(limitesPear3$violations$lower)
alpha_3[i] <- alphaPear3/n

```

```

limitesPear4 <- cusum(residPear_out,center=mean(residPear_in),
std.dev=sd(residPear_in), decision.interval=3*2.050763,
se.shift=1,plot=FALSE)
alphaPear4 <- length(limitesPear4$violations$upper)+
length(limitesPear4$violations$lower)
alpha_4[i] <- alphaPear4/n

limitesPear5 <- cusum(residPear_out,center=mean(residPear_in),
std.dev=sd(residPear_in), decision.interval=3*2.06,
se.shift=1,plot=FALSE)
alphaPear5 <- length(limitesPear5$violations$upper)+
length(limitesPear5$violations$lower)
alpha_5[i] <- alphaPear5/n

limitesPear6 <- cusum(residPear_out,center=mean(residPear_in),
std.dev=sd(residPear_in), decision.interval=3*2.07,
se.shift=1,plot=FALSE)
alphaPear6 <- length(limitesPear6$violations$upper)+
length(limitesPear6$violations$lower)
alpha_6[i] <- alphaPear6/n

}

# CUSUM - METODO 1
arl1 <- 1/mean(alpha_1)
arl2 <- 1/mean(alpha_2)
arl3 <- 1/mean(alpha_3)
arl4 <- 1/mean(alpha_4)
arl5 <- 1/mean(alpha_5)
arl6 <- 1/mean(alpha_6)
mrl1<-log(0.5)/log(1-mean(alpha_1))
mrl2<-log(0.5)/log(1-mean(alpha_2))
mrl3<-log(0.5)/log(1-mean(alpha_3))
mrl4<-log(0.5)/log(1-mean(alpha_4))
mrl5<-log(0.5)/log(1-mean(alpha_5))
mrl6<-log(0.5)/log(1-mean(alpha_6))
sdr11<-sqrt((1-mean(alpha_1))/(mean(alpha_1)^2))

```

```

sdrl2<-sqrt((1-mean(alpha_2))/(mean(alpha_2)^2))
sdrl3<-sqrt((1-mean(alpha_3))/(mean(alpha_3)^2))
sdrl4<-sqrt((1-mean(alpha_4))/(mean(alpha_4)^2))
sdrl5<-sqrt((1-mean(alpha_5))/(mean(alpha_5)^2))
sdrl6<-sqrt((1-mean(alpha_6))/(mean(alpha_6)^2))

mresults <- rbind(mresults,c(delta,arl1,arl2,arl3,
arl4,arl5,arl6,mrl1,mrl2,mrl3,mrl4,mrl5,mrl6,sdrl1,
sdrl2,sdrl3,sdrl4,sdrl5,sdrl6))

}

(proc.time()-t0)/60

# METODO 1 -----
colnames(mresults)<-c("Delta","IBRCC-ARL","Pearson4-ARL","IBRCC-MRL",
"Pearson4-MRL", "IBRCC-SDRL","Pearson4-SDRL")

print(mresults)

w1 <- seq(2.04,2.07,0.01)
Pearson_ARL <- as.numeric(mresults[6,c(3:7)])
larl1 <- log(Pearson_ARL)
fit1 <- lm(w1 ~ larl1)
myfit <- function(ARL,fit)
{
  sum((coef(fit))*c(1,log(ARL)))
}
round(myfit(100,fit1),6)

```