UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Estimadores diretos para índices de capacidade em um processo em tempo real para tomada de decisão em controle de qualidade estatístico

Ana Paula Silva Figueiredo

Orientador: Prof. Dr. Fernando Antônio Moala

Coorientador: Prof. Dr. Pedro Luiz Ramos

Programa: Matemática Aplicada e Computacional

Presidente Prudente, Agosto de 2022

UNIVERSIDADE ESTADUAL PAULISTA

Faculdade de Ciências e Tecnologia de Presidente Prudente

Programa de Pós-Graduação em Matemática Aplicada e Computacional

Estimadores diretos para índices de capacidade em um processo em tempo real para tomada de decisão em controle de qualidade estatístico

Ana Paula Silva Figueiredo

Orientador: Prof. Dr. Fernando Antônio Moala

Coorientador: Prof. Dr. Pedro Luiz Ramos

Dissertação apresentada ao Programa de Pós-Graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP para obtenção do título de Mestre em Matemática Aplicada e Computacional.

Presidente Prudente, Agosto de 2022

F475e

Figueiredo, Ana Paula Silva
Estimadores diretos para índices de capacidade em um processo em tempo real para tomada de decisão em controle de qualidade estatístico / Ana Paula Silva Figueiredo. -- Presidente Prudente, 2022
76 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente
Orientador: Fernando Antonio Moala
Coorientador: Pedro Luiz Ramos

1. Estatística. 2. Controle de qualidade estatístico. 3. Índices de capacidade. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Presidente Prudente

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Estimadores diretos para índices de capacidade em um processo em tempo real para tomada de decisão em controle de qualidade

AUTORA: ANA PAULA SILVA FIGUEIREDO

ORIENTADOR: FERNANDO ANTONIO MOALA

Aprovada como parte das exigências para obtenção do Título de Mestra em Matemática Aplicada e Computacional, pela Comissão Examinadora:

Prof. Dr. FERNANDO ANTONIO MOALA (Participaçao Virtual)
Departamento de Estatistica / Faculdade de Ciencias e Tecnologia de Presidente Prudente

Prof. Dr. PEDRO LUIZ RAMOS (Participaçao Virtual)
Facultad de Matematicas / Pontificia Universidad Católica de Chile

Prof. Dr. EDILSON FERREIRA FLORES (Participaçao Virtual)
Departamento de Estatística / Faculdade de Ciencias e Tecnologia de Presidente Prudente

Presidente Prudente, 31 de agosto de 2022

Dedico este trabalho a todos professores que tive desde a primeira infância até os dias de hoje, a minha família e amigos por sempre me apoiarem em ir atrás dos meus sonhos.

Agradecimentos

Agradeço primeiramente a Deus pela oportunidade de chegar até aqui, de conhecer pessoas e possibilidades completamente diferentes do que eu imaginei.

Gostaria de agradecer a todos que de alguma forma fizeram com que eu trilhasse este caminho, desde os professores do ensino fundamental até os da graduação e pós - graduação, amigos e família.

Agradecer também ao meu orientador Dr. Professor Fernando Antônio Moala e ao Prof. Dr. Pedro Luiz Ramos, pela paciência e disponibilidade. Agradecer ao Prof. Sérgio Minoru Oikawa pelo apoio deste os primeiros anos do curso de graduação.

Aos meus amigos mais próximos, que apesar da pandemia e rumos diferentes que tomamos fazem parte de quem sou. À minha mãe que me ensinou que a vida é aqui e agora, e que somos nós os responsáveis pelas escolhas que fazemos. Agradecer ao César Augusto, meu namorado, pela cumplicidade, paciência e apoio de sempre.

Agradecer também ao Fernando Pacanelli Martins, técnico responsável pelo cluster, sempre prestativo e atencioso aos problemas enfrentados por mim no uso da ferramenta. Desta forma, registro que os resultados gerados neste trabalho foram produzidos com o uso do Cluster do Laboratório de Simulação Numérica da FCT/Unesp.

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

"Hoje me sinto mais forte
Mais feliz, quem sabe
Só levo a certeza
De que muito pouco sei
Ou nada sei"
Renato Teixeira

Resumo

O avanço da tecnologia fomentou a competitividade em todos nichos, e no ramo de manufatura não foi diferente. Desta forma, tornou-se necessário a atuação de órgãos que regulem e estabeleçam normas para o processo de confecção dos produtos com objetivo de formalizar um padrão de qualidade e também proteger o consumidor final. Por isto, a avaliação constante de processos é indispensável. Os índices de capacidade são ferramentas que indicam a situação do processo, apontando se ele é capaz ou não de atentar os limites de especificações previamente definidas pelos órgãos reguladores. O objetivo deste trabalho é verificar a situação de um processo, em tempo real, através de uma ferramenta gráfica. O gráfico utiliza o índice de capacidade C_{pk} para classificar o processo como satisfatório ou insatisfatório. Resumidamente, propomos uma ferramenta gráfica que indique a situação de um processo, utilizando um índice de capacidade estimado através de amostras de um processo em tempo real, de forma rápida e direta, utilizando assim estimadores em forma fechada, para auxiliar em tomadas de decisões.

Palavras-Chave: *Processo em Tempo Real; Índice C_{pk}; Estimadores em forma fechada; Ferramenta gráfica.*

Abstract

The advancement of technology has fostered competitiveness in all niches, and in the manufacturing industry it was no different. Thus, it has become necessary the action of agencies that regulate and establish standards for the process of making products with the aim of formalizing a standard of quality and also protect the final consumer. Therefore, the constant evaluation of a process is indispensable. The capability indexes are tools that indicate the situation of the process, pointing out whether it is able or not to meet the limits of specifications pre-defined by the regulatory agencies. The objective of this work is to verify the situation of a process, in real time, through a graphic tool. The graph uses the capability index C_{pk} to classify the process as satisfactory or unsatisfactory. In summary, we propose a graphical tool that indicates the situation of a process, using a capability index estimated through samples of a process in real time, in a fast and direct way, thus using closed-form estimators, to aid in decision making.

Keywords: *Real-Time Process; Index C_{pk}; Estimators in closed-form; Graphical tool.*

Lista de Figuras

Lista de Tabelas

Lista de Siglas

MV: Método de maxíma verrosimilhança.

MVM: Método de maxíma verrosimilhança modificado.

MLM: Método L-Momentos.

ICP: Índices de capacidade de processo.

LIC: Limite inferior de controle.

LC: Limite de controle central.

LSC: Limite superior de controle.

LSE: Limite superior de especificação.

LIE: Limite inferior de especificação.

FDP: Função densidade de probabilidade.

FDA: Função distribuição acumulada.

VP: Verdadeiro Positivo.

VN: Verdadeiro Negativo.

FP: Falso Positivo.

FN: Falso Negativo.

$DP_{C_{pk}}$: Desvio Padrão do índice C_{pk}.

RME: Erro médio relativo.

EQM: Erro quadrático médio.

iid: Independente Identicamente Distribuída.

INMETRO: Instituto Nacional de Metrologia, Qualidade e Tecnologia.

Sumário

Capítulo

1

Introdução

A evolução tecnológica, exigência por um padrão de qualidade, tanto por parte dos órgãos regulamentadores como pelo mercado consumidor, trouxe a necessidade de avaliação constante durante todos os processos de fabricações nas industrias e fábricas. Sendo assim, propostas de ferramentas que auxiliam a tomada de decisões, referentes a situação do processo são altamente necessárias e bem vindas

Uma forma de medir a capacidade do processo é através dos índices de capacidades de processos (ICP's), que são métricas utilizadas em análise de capacidade de processo com objetivo de avaliar o estado da manufatura, ou seja, se é capaz ou não de atender as especificações dadas. A ideia principal é otimizar a produção, reduzindo assim itens obsoletos, como visto em Spiring *et al.* [22]. A literatura apresenta vários tipos de ICP's, cada um com sua particularidade, mas todos são intencionados em checar a situação do processo. O mais utilizado atualmente é o índice C_{pk}, que considera também a média como a variação presentes no processo, por este motivo é o utilizado neste trabalho.

Inicialmente, os índices propostos assumiam que os dados seguiam distribuição normal, porém em algumas situações está condição não era satisfeita o que incentivou alguns autores a proporem algumas soluções como por exemplo a transformação dos dados deixando-os próximos a distribuição normal, e após isto estimar os ICP's. Podemos citar o sistema de transformação de Johnson [9], também o método de transformação de Box-Cox [1], e entre vários outros que podem ser vistos em Rivera *et al.* [18], Somerville e Montgomery [21], Pinã-Monarrez *et. al* [16].

Outra alternativa para lidar com dados não gaussianos é ajustar uma distribuição apropriada ao problema e após calcular os novos índices usando as medidas da distribuições em questão. Como Clements [2] que propôs o método de percentis para calcular os ICP's, Kotz e Lovelace [10] que construíram tabelas baseadas em caudas padronizadas das curvas de Pearson como funções de curtose e assimetria, Housseinifard *et al.* [8] que discutiram diferentes métodos para se estimar os ICP's para um processo não gaussiano.

Vale ressaltar que para aplicar-se as métricas ICP's é necessário que o processo esteja sob controle estatístico, ou seja, o processo deverá estar estável, para isto, uma alternativa, é utilizar as cartas de controle de Shewhart. De modo geral, elas monitoram o processo e indicam, no decorrer do tempo, se ele é estável ou se depende de fatores externos inerentes a produção.

Uma distribuição que descreve vários fenômenos e é a base para muitas teorias estatísticas é a distribuição Normal. Como visto, é um dos pressupostos assumidos para calcular os primeiros índices de capacidade, uma vez que todos processos existentes na época eram descritos por ela. Estimar seus parâmetros pelo método de verossimilhança resulta em estimativas diretas, não sendo necessário artifícios numéricos para as encontrar. Sendo

assim, neste trabalho assumi-se o método de verossimilhança como forma de estimação direta para a Distribuição Normal.

Outra distribuição amplamente utilizada no controle de qualidade é a distribuição Gama, que carrega boas propriedades para o método de estimação de verossimilhança como eficiência, consistência e invariância para transformações um-para-um e também sendo o caso generalizado de distribuições como Exponencial e Qui-Quadrado. Embora possua as vantagens citadas não apresenta uma função de estimação na forma fechada, para obtenção dos parâmetros em estudo, sendo necessário o uso de métodos numéricos para encontrar os estimadores, gerando um custo computacional considerável.

Desta forma, Ramos *el al.* [17] propuseram uma modificação no método de verossimilhança, denominado método de verossimilhança modificado (MVM), que consiste em usar uma distribuição que contenha um parâmetro extra, que ao ser fixado se torna um caso particular do modelo alvo, ela é usada para encontrar os estimadores de forma explícita. No caso da distribuição Gama é usada a distribuição Gama Generalizada, que possui três parâmetros, sendo possível aplicar o método e obter a função de estimação na forma direta.

Agora, destacamos também a distribuição Weibull que também é utilizada em diversas áreas do conhecimento para modelar os mais variados tipos de dados, sendo casos gerais de distribuições como por exemplo as distribuições Exponencial e Rayleigh. Embora a estimação de verossimilhança para a distribuição Weibull seja assintoticamente normal e eficiente para amostras com tamanho grande, também não possui expressões em forma fechada para os estimadores, sendo necessário o uso de métodos iterativos elevando o custo computacional.

Sendo assim, Hosking [7] trabalhou no método L - Momentos, que faz uso dos l-momentos amostrais e populacionais para obtenção dos estimadores. Em suas aplicações o método mostrou-se robusto na presença de outliers e eficiente mesmo para amostras pequenas, como mostrado em Elamir e Allan [5] e Teimouri *et al* [23], e para a distribuição Weibull este método produz estimativas diretas.

Atualmente, não há ferramentas gráficas que incluam a estimação do índice C_{pk} para um processo em tempo real. Desta forma, a motivação deste trabalho é criar uma ferramenta gráfica, utilizando o índice C_{pk} apresentado por Clements [2] que seja capaz de verificar as mudanças no processo de forma visual, em tempo real, utilizando os métodos de estimação apresentados anteriormente. Para isto, criamos uma função no software livre R, que recebe dados do usuário, juntamente com a distribuição que tais dados pertencem, e retornam um gráfico e uma tabela com os valores pontuais e seus respectivos intervalos de 95% de confiança. No gráfico há faixas que indicam a situação do processo em cada amostra retirada, ou seja, se o processo é altamente capaz, moderadamente capaz ou incapaz.

Considerando a justificativa de diminuir o custo computacional ao usarmos os estimadores de forma direta para as diferentes distribuições, um estudo de simulação foi realizado para quantificar as proporções de acertos para ambos tipos de simulação: direto e indireto. E além disto, as simulações foram realizadas para valores pontuais e intervalares do índice C_{pk}.

Destacamos também que, o estudo de simulação foi realizado para três cenários diferentes, chamados de fronteira, ótimo e descontrolado. O primeiro cenário leva em consideração um índice C_{pk} próximo de $1,01$, este cenário está na fronteira entre o que é considerado capaz ou não. Para o ótimo, o valor do índice estimado é em torno de $1,33$, pois para valores maiores ou iguais a este o processo é considerado altamente capaz e por fim, o cenário descontrolado com índice em torno de $0,70$, que pela interpretação o processo é considerado incapaz de atender as especificações, uma vez que é menor que

um. O objetivo de separar estes três cenários é verificar qual a performance do modelo em diferentes "ambientes".

Deste modo, o relatório está organizado na seguinte forma: no capítulo 2 é apresentado uma breve revisão sobre as cartas de controle e os índices de capacidade de processos, sobre o método de estimação de verossimilhança, a reamostragem de bootstrap e também os modelos de distribuições utilizados que são a Normal, Gama e Weibull e suas características. E ainda, nos capítulos 3 e 4, mostra-se os métodos de estimações de forma direta para as distribuições Gama e Weibull, no caso, o Método de Verossimilhança Modificado (MVM) e o Método L-Momentos (MLM), respectivamente. No capítulo 5 é apresentado a proposta gráfica para a análise do índice no decorrer do tempo. E ainda no capítulo 6, o estudo de simulação para processos em tempo real. E por fim, nos capítulos 7, 8, uma aplicação e considerações finais do método exposto.

Capítulo

2

Revisão

Neste capítulo é exposto uma revisão sobre algumas ferramentas de controle estatístico de qualidade, índices de capacidade, o método de estimação de verossimilhança e as distribuições usadas neste trabalho.

2.1 Cartas de Controle

A necessidade em padronizar processos, diminuir a variabilidade e refugos, aumentar a qualidade e, consequentemente, os ganhos de um processo produtivo incentivou diversos estudiosos a criarem ferramentas que avaliassem os processos.

Desta forma, na década de 1920 o físico Doutor Walter Shewhart iniciou a teoria do controle estatístico de qualidade, que se utiliza de ferramentas que mostram visualmente a variabilidade de um processo. De modo geral verificam o processo afim de estabiliza-lo ao decorrer do tempo, resultando em um produto final com melhor qualidade, menor variação e também otimizar os recursos disponíveis usados.

As ferramentas do controle estatístico de processo são as cartas de controle, conhecidos também como gráficos de Shewhart.Resumidamente, existem as cartas de controles de atributos e das variáveis, neste trabalharemos com as cartas para as variáveis. O objetivo das cartas para as variáveis é monitorar um processo verificando sua média e variabilidade, para então indicar se o processo é ou não estável. De modo geral, os gráficos de Shewhart verificam determinado parâmetro θ_X, seja ele a média ou variância do processo de interesse, e indicam se eles estão constantes durante um período de tempo ou se depende de fatores externos durante sua produção.

Sendo assim, analisa-se a partir de uma amostra de uma variável aleatória X o desempenho do parâmetro θ_X pela sua estimativa entre os limites de controle da carta, que são eles: o limite de controle inferior (LCI), limite de controle central (LC) e o limite de controle superior (LCS).

E ainda, ao assumirmos uma política de qualidade seis sigma, e se a distribuição de X puder ser aproximada por uma distribuição Normal com uma média μ_X e desvio padrão σ_X, então os limites inferior e superior para o parâmetro θ_X são dados por:

$$\mu_X \pm 3\sigma_X. \tag{2.1}$$

Nesta seção verificaremos apenas dois destes gráficos, a carta para média e para a amplitude nas subseções 2.1.1 e 2.1.2, respectivamente.

2.1.1 Carta para Média

Conhecida também por carta $\overline{X}$, a carta para média verifica como estão distribuídas as médias do processo.

Seja X uma variável com determinada distribuição de probabilidade cujo parâmetros para média e desvio padrão são μ e σ, respectivamente. Sabemos que a média amostral, desta variável, obtida de uma amostra de tamanho n é dada por $\overline{X} = \sum_{i=1}^{n} \frac{x_i}{n}$.

Suponhamos um processo no qual há m subamostras com tamanhos iguais a n, então a média amostral deste processo será a média das médias, que representa a linha de controle central (LC) da carta $\overline{X}$ ou seja :

$$LC_{\bar{X}} = \overline{\overline{X}} = \sum_{j=1}^{m} \frac{\overline{X}_j}{m} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}. \tag{2.2}$$

Seguindo (2.1), temos que os limites de controle para a carta da média são dados por:

$$\begin{aligned} LCI_{\bar{X}} &= \overline{\overline{X}} - 3\frac{\hat{\sigma}}{\sqrt{n}}, \\ LCS_{\bar{X}} &= \overline{\overline{X}} + 3\frac{\hat{\sigma}}{\sqrt{n}}, \end{aligned} \tag{2.3}$$

sendo $LCI_{\bar{X}}$ e $LCS_{\bar{X}}$ os limites de controle inferior e superior da carta $\overline{X}$, respectivamente.

E ainda, assumindo que a variável possa ser aproximada da distribuição normal, então uma estimativa para σ é dado por:

$$\hat{\sigma} = \frac{\overline{R}}{d_2}, \tag{2.4}$$

com $\overline{R}$ a amplitude média e d_2 fator de correção para tamanho da subamostra apresentado em Apêndice C na Tabela C.5.

2.1.2 Carta para amplitude

A carta R ou a carta para amplitude é uma ferramenta utilizada para verificar a variabilidade de um processo em que há poucas observações nas subamostras, menos de 10 ou 12, segundo Louzada *et. al.* [11], e ainda todas possuem o mesmo tamanho.

Deste modo, a amplitude amostral da *i-ésima* subamostra é dada pela por:

$$R_i = max_j(x_{ij}) - min_j(x_{ij}),$$

com $i = 1, 2, ..., m$ e $j = 1, 2, ..., n$.

Sendo assim, o limite de controle central para a carta R é a média das amplitudes dada por:

$$LC_R = \sum_{i=1}^{m} \frac{R_i}{m} = \overline{R}. \tag{2.5}$$

A estimativa para o desvio padrão da amplitude é dado por $\hat{\sigma_R} = d_3\hat{\sigma}$, sendo d_3 o fator de correção amostral e $\hat{\sigma}$ dado em (2.4). Temos, enfim, os limites de controle inferior e superior para a carta R:

$$\begin{aligned} LCI_R &= \overline{R} - 3d_3\frac{\overline{R}}{d_2}, \\ LCS_R &= \overline{R} + 3d_3\frac{\overline{R}}{d_2}. \end{aligned} \tag{2.6}$$

Lembrando que os valores dos fatores de correção estão descritos na Tabela C.5.

2.1.3 Principais regras sensibilizantes

Vale ressaltaar que mesmo que todos os pontos estejam dentro dos limites de controle não significa que o processo está estável. Conforme cita Louzada *et. al.* [11]: a não aleatoriedade, padrão de repetição, seja cíclico ou sazonal, e também quebras de regras sensibilizantes, podem indicar um processo não estável.

Segundo Montgomery [15], as regras sensibilizantes mais utilizadas no território prático são:

- Um ou mais pontos fora dos limites de controle;
- Dois ou três pontos seguidos do mesmo lado do limite $(\mu + 2\sigma)$ ou além;
- Quatro ou cinco pontos consecutivos do mesmo lado do limite em $(\mu - 2\sigma)$ ou além;
- Oito pontos seguidos de um mesmo lado da linha central;
- Seis pontos seguidos em ordem crescente ou decrescente.

Os limites $\mu \pm 2\sigma$ são chamados de limites de alerta.

2.2 Índices de Capacidade

Os índices de capacidade são ferramentas utilizadas para avaliar o desempenho de um processo fabril com o objetivo de produzir itens dentro das especificações predefinidas, diminuindo assim a variabilidade no processo de fabricação, aumentando seu potencial de ganho e sua força competitiva com outras empresas.

Para estimar-se a capacidade de um processo é preciso quantificar uma medida que explique seu desempenho de fabricação. Através desta análise é possível dizer em relação a qualidade dos itens produzidos, se estão ou não dentro dos limites de especificação, que são definidos pelos órgãos regulamentadores.

Inicialmente, as inspeções de qualidade tinham o objetivo de verificar se a produção estava ou não estável, através das cartas de controle, e também apoiar ações que visavam aumentar a qualidade do produto. Nesse sentido, os ICP's são ferramentas importantes, uma vez que possuem fortes interpretações e também são facilmente calculados. Porém para serem estimados algumas suposições devem ser satisfeitas para que se garanta a integridade dos resultados. As suposições são:

- Os índices devem ser calculados em amostras que possuam parâmetros controlados, ou seja, que estejam sob controle estatístico de qualidade;
- A variável observada deve seguir a distribuição normal;
- A média se não corresponder ao centro dos limites de especificação, LIE (Limite Inferior de Especificação) e LSE (Limite Superior de Especificação), deve ao menos estar entre eles.

Segundo Louzada *et al.* [13], se as observações não seguem uma distribuição Normal, é necessário transforma-las para que seja possível aproxima-las à normalidade, uma alternativa é a transformação de Box-Cox, que na prática não é muito indicada uma vez que a interpretação é realizada com os dados transformados. É possível também, utilizar a abordagem quantílica para medir a estimativa paramétrica dos índices que foi abordada por Clements [2], e que usaremos nesse.

Muitos autores como Pinã-Monarrez *et al.* [16], Sennaroglu e Senvar [19], Senvar *et al.* [20], Hosseinifard *et al.* [8] mostram estudos nos quais surgem problemas quando se realiza a análise de capacidade do processo assumindo normalidade de dados quando, na verdade, não há. Isto leva a conclusões equivocadas sobre o processo, uma vez que os índices não são precisos.

Desta forma, neste capítulo, nas próximas subseções, descrevemos alguns índices utilizados para dados que apresentam distribuição Normal e também para aqueles que não são normalmente distribuídos.

2.2.1 Índice para um processo Gaussiano

Nesta seção apresenta-se alguns índices de capacidade de processos utilizados quando os dados são modelados segundo a distribuição normal.

2.2.1.1 Índice C_p

O mais simples dos índices, o C_p é dado por:

$$C_p = \frac{LSE - LIE}{6\sigma}, \tag{2.7}$$

em que σ é o desvio padrão populacional do processo.

Ele compara a variabilidade natural do processo com a tolerável. É utilizado para processos centrados na média. Valores de $C_p < 1$ dão indícios de que o processo não dispõe de condições para atender as especificações preestabelecidas. Como nem sempre é possível conhecer a variabilidade populacional do processo, então estima-se o índice, que é dado por:

$$\hat{C}_p = \frac{LSE - LIE}{6S}, \tag{2.8}$$

com $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$, o desvio padrão amostral do processo, e n a quantidade de itens observado na amostra em questão.

Observe que o C_p não considera a posição do processo, então não é possível avaliar se ele está dentro ou ao menos perto da região definida pelos limites de especificações.

2.2.1.2 Índice C_{pk}

Como visto anteriormente, índice C_p não diz sobre a posição do processo, desta forma o índice C_{pk} é uma alternativa, pois é uma combinação do anterior que considera a localização e a variabilidade do processo. Útil para processos não centrados na média, pois considera tanto a média quanto variância de tal. Apresentado por:

$$C_{pk} = min\Big(\frac{LSE - \mu}{3\sigma}, \frac{LIE - \mu}{3\sigma}\Big), \tag{2.9}$$

sendo LSE e LIE os limites superior e inferior de especificação, respectivamente, μ e σ a média e o desvio padrão populacionais dos dados.

Valores de $C_{pk} < 1$ indicam um processo incapaz de atender as especificações, para $1 \leq C_{pk} < 1,33$ o processo é considerado capaz mas com uma certa cautela, já para valores de $C_{pk} \geq 1,33$ o processo é considerado altamente capaz de atentar as especificações impostas.

Observação 1 *Se $C_p = C_{pk}$ significa que o processo é centrado em torno da média da especificação, por outro lado se $C_p \neq C_{pk}$ então a média do processo não é a mesma dos*

limites de especificações fornecidos, ou seja, o processo não está centrado em torno da média.

2.2.1.3 Índice C_{pm}

O índice C_{pm} é dado por meio de uma razão do índice C_p e a raiz quadrada envolvendo a média do processo, a média da especificação nominal e a variância do processo. Que é obtido como:

$$C_{pm} = \frac{LSE - LIE}{6\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{C_p}{\sqrt{1 + (\mu - T)^2}}, \tag{2.10}$$

sendo T : o ponto médio do intervalo de especificação.

Se o processo apresenta uma grande variabilidade, o denominador irá aumentar, causando uma diminuição do valor de C_{pm}. E também se a distância entre T e μ for considerável, pode diminuir o valor do índice.

Observação 2 *É importante notar que se $\mu = T$ temos $C_{pm} = C_p$.*

2.2.2 Índices para um processo não gaussiano

Como visto anteriormente, quando os dados são não-normais é um erro aplicar os índices já mencionados pois a interpretação da métrica estará equivoca, já que os índices são construídos baseados sob suposição de normalidade da amostra.

Desta forma, Clements [2] apresentou um método alternativo para realizar a análise de capacidade do processo para dados com distribuições não-normais. Aqui mostram-se índices semelhantes aos anteriores mas agora assumindo a distribuição quantílica dos dados.

2.2.2.1 Índice C_p para distribuição não normal

O índice C_p para distribuição não normal, denotado por C_p^*, como o anterior considera apenas a dispersão do processo e não sua localização em relação aos limites de especificação. Sendo assim, ele é definido como:

$$C_p^* = \frac{LSE - LIE}{L_p - U_p}, \tag{2.11}$$

em que L_p e U_p são os percentis 99,865 e 0,135, respectivamente, da distribuição dos dados.

2.2.2.2 Índice C_{pk} para distribuição não normal

Já pontuamos que o índice anterior não é capaz de analisar sobre a posição do processo em relação aos limites de especificações. Assim, apresentou-se o índice C_{pk}^* para distribuição não normal, que considera tanto a variabilidade do processo e também sua localização. Ele é calculado através de:

$$C_{pk}^* = min\left(\frac{LSE - M}{U_p - M}, \frac{M - LIE}{M - L_p}\right), \tag{2.12}$$

sendo M, L_p e U_p os percentis de números 50; 99,865 e 0,135, respectivamente, da distribuição utilizada.

2.2.2.3 Índice C_{pm} para distribuição não normal

O Índice C^*_{pm} para distribuição não normal é semelhante ao C_{pm} visto na seção anterior. É dado por:

$$C^*_{pm} = \frac{LSE - LIE}{6\sqrt{\left[L_p - F_{0,135}\right]^2 + (M-T)^2}}, \tag{2.13}$$

onde L_p e U_p são os percentis 99,865 e 0,135 da distribuição em questão, M é a mediana do processo, e T valor nominal das especificações dos órgãos reguladores.

2.3 Método de Máxima Verossimilhança

O estimador de máxima verossimilhança é baseado nos resultados obtidos a partir da amostra. Ele define os valores dos parâmetros que melhor expliquem a amostra utilizada, ou seja, estima valores para cada parâmetro baseado nos dados empíricos. Abaixo, é apresentado a definição da função de verossimilhança dada por Bolfarine e Sandoval [3].

Definição 1 *Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de tamanho n da variável aleatória X com função densidade (ou de probabilidade) $f(\boldsymbol{x}|\boldsymbol{\theta})$ com $\boldsymbol{\theta} \in \Theta$, onde Θ é o espaço paramétrico. A função de verossimilhança de $\boldsymbol{\theta}$ correspondente à amostra aleatória observada é dada por*

$$L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}). \tag{2.14}$$

Os estimadores são obtidos a partir dos seguintes passos:

1. Construir a função de verossimilhança, dada na equação (2.14);
2. Linearizar a função, encontrada através do logaritmo da verossimlhança:

$$log(L(\boldsymbol{\theta}; \boldsymbol{x})) = l(\boldsymbol{\theta}; \boldsymbol{x});$$

3. Igualar as derivas parciais a zero, ou seja:

$$\frac{\partial}{\partial \theta_i} l(\boldsymbol{\theta}; \boldsymbol{x}) = 0, i = 1, 2, ..., p. \tag{2.15}$$

Muitas vezes, o sistema resultante da equação (2.15) não é linear, e métodos numéricos são necessários para encontrar os estimadores de máxima verossimilhança, um deles é o método de Newton-Raphson, descrito na seção 2.3.2.

2.3.1 Propriedades do Estimador de Verossimilhança

Algumas propriedades apresentadas por Meier [14] e Bolfarine e Sandoval [3] são:

- Invariância: se $\hat{\theta}$ é uma estimativa de MV, então a estimativa para função $g(\theta)$, sendo g monótoma contínua é $g(\hat{\theta})$;
- A estimativa de MV é uma estatística suficiente, ou seja, o estimador descreve a amostra sem perda de informação;

- Para amostras grandes os estimadores de verossimilhança são assintoticamente normais, não-viciados e eficientes, ou seja:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim^{n\to\infty} \boldsymbol{N_p}\left[\mathbf{0}, \boldsymbol{I^{-1}(\theta)}\right], \tag{2.16}$$

com $I(\boldsymbol{\theta})$ a matriz de informação de Fisher, sendo cada item formado por:

$$\{I_{ij}(\boldsymbol{\theta})\} = \left\{E\left[\left(\frac{\partial}{\partial\theta_i\theta_j} log f(\boldsymbol{x}|\boldsymbol{\theta})\right)^2\right]\right\}, \quad i,j = 1,2,...,p. \tag{2.17}$$

2.3.2 Método Numérico de Newton-Raphson

Criado simultaneamente por Isaac Newton e Joseph Raphson, o método Newton-Raphson é um processo iterativo, que a partir de um valor inicial obtêm-se uma sequência de pontos que convergirá para a raiz da função.

Encontrar-se a solução da equação (2.15), pelo processo iterativo de Newton-Raphson que é dado por:

$$\boldsymbol{\theta^{(n+1)}} = \boldsymbol{\theta^{(n)}} - \left[J\left(\boldsymbol{\theta^{(n)}}\right)\right]^{-1} D(\boldsymbol{\theta}), \tag{2.18}$$

sendo $D\left(\boldsymbol{\theta}\right)$ um vetor coluna das derivadas parciais apresentado pela equação (2.15) e $\left[J\left(\boldsymbol{\theta^{(n)}}\right)\right]^{-1}$ a inversa da matriz Jacobiana para as derivadas apresentadas na mesma equação.

2.4 Reamostragem de Bootstrap

O método de reamostragem de Boostrap, neste trabalho, foi utilizado como alternativa para criar os intervalos de confiança para as distribuições assintóticas desconhecidas dos índices gerados.

Proposto por Efron [4], o método de reamostragem Boostrap consiste em re-amostrar uma amostra observada, e a partir disto estimar a variabilidade da medida em si. Neste, em particular, o objetivo é estimar um intervalo de boostrap paramétrico para o índice C_{pk}.

Seja $(Y_1, Y_2, ..., Y_n)$ uma amostra aleatória de tamanho n de $F(y|\theta)$. Deseja-se estimar o índice C_{pk} a partir de $\hat{C}_{pk} = T(y)$. Sendo assim, os passos para chegar-se nos intervalos para a medida foram:

- Obter a amostra $F(y|\hat{\boldsymbol{\theta}})$;
- Reamostrar B vezes a amostra $Y^* = (Y_1^*, Y_2^*, ..., Y_n^*)$ de $F(y|\hat{\boldsymbol{\theta}})$;
- Calcular em cada reamostra a estimativa para $\hat{C}_{pk}$.

Assim, o intervalo de 95% de confiança paramétrico para o $\hat{C}_{pk}$ é:

$$I.C_{boot} = \left[\hat{C}_{pk} - q_{0,025}EP(\hat{C_{pk}}), \hat{C}_{pk} + q_{0,975}EP(\hat{C_{pk}})\right]. \tag{2.19}$$

No qual $EP(\hat{C_{pk}}) = \frac{DP_{\hat{C_{pk}}}}{\sqrt{n}}$ é o erro padrão de $\hat{C_{pk}}$ e q_α quantil de α% da re-amostra.

2.5 Distribuições de Probabilidade

Nesta seção as características das distribuições Normal, Gama e Weibull, usadas no trabalho, são apresentadas. Expomos suas funções de distribuições acumulada (FDA), de densidade de probabilidade (FDP), de sobrevivência e também da função risco. A primeira subseção é dedicada a distribuição Normal, após para distribuição Gama, e por fim para distribuição Weibull.

2.5.1 Distribuição Normal

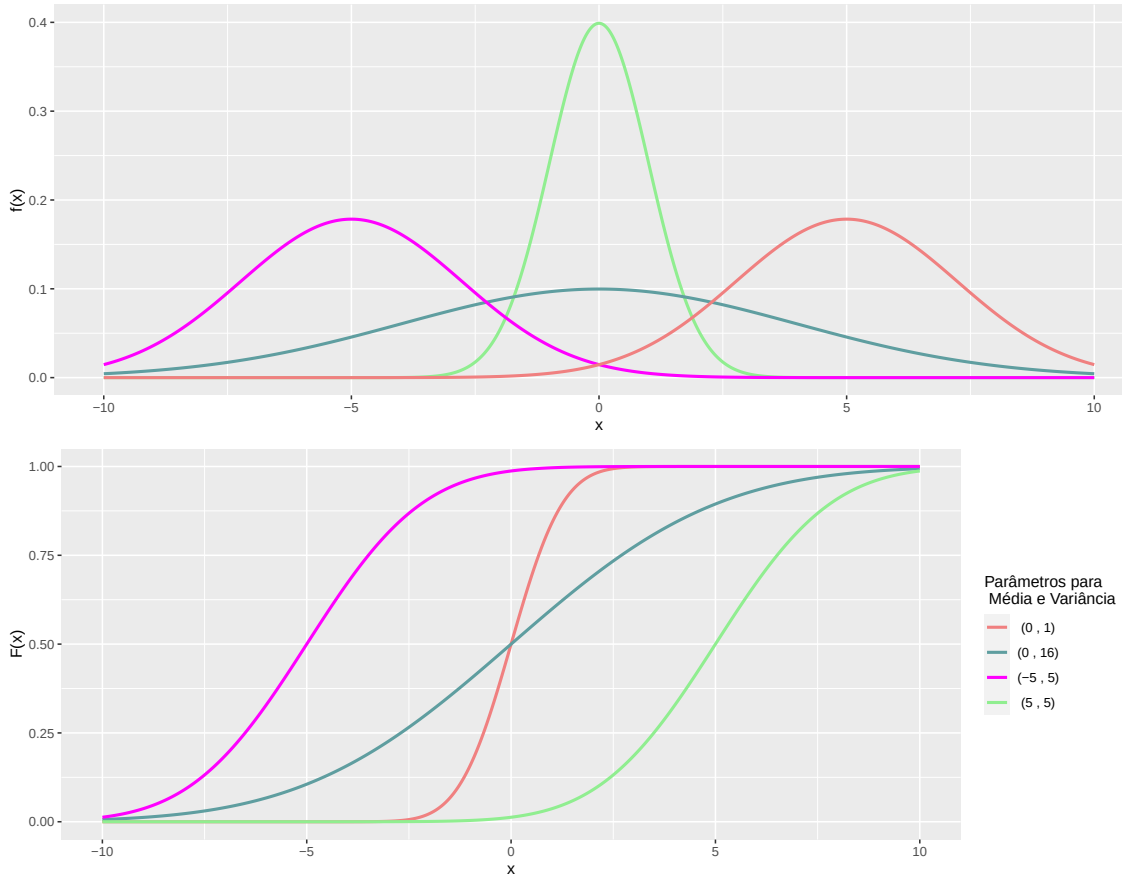
A distribuição Normal, base da teoria estatística, que descreve vários fenômenos de medidas é apresentada nesta subseção. Seja Y uma variável aleatória contínua segundo a distribuição Normal com uma média μ e variância σ^2, então suas funções de densidade de probabilidade e de distribuição acumulada são definidas por:

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} \quad e \quad F(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx, \tag{2.20}$$

com $-\infty < y < \infty$, $-\infty < \mu < \infty$ e $\sigma^2 > 0$.

Na Figura 2.1 temos os gráficos das funções mencionadas em 2.20 para pares de parâmetros diferentes.

Figura 2.1: Função de distribuição de probabilidade (A) e Função densidade Acumulada (B) da distribuição Normal.



Observamos que a média, moda e mediana da distribuição normal compartilham do mesmo valor, no caso, μ.

2.5.1.1 Algumas Medidas da Distribuição Normal

Nesta subseção, algumas medidas da distribuição normal são apresentadas.

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\Big\{ -\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\} dx. \tag{2.21}$$

Fazendo $u = \frac{(x-\mu)}{\sigma}$, temos da equação 2.21:

$$\frac{1}{\sqrt{2\pi}}\left[\sigma \int_{-\infty}^{\infty} u \exp\Big\{ -\frac{u^2}{2}\Big\} du + \mu \int_{-\infty}^{\infty} \exp\Big\{ -\frac{u^2}{2}\Big\} du\right].$$

Observe que a função na primeira integral é uma função ímpar, logo o resultado desta é zero. Temos então:

$$\mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{u^2}{2}\Big\} du,$$

note que $U \sim$ Normal $(0, 1)$, portanto sua integral é um. Desta forma:

$$E(X) = \mu. \tag{2.22}$$

Calculemos agora a variância para a distribuição. Sabemos que $Var(X) = E(X^2) - (E(X))^2$, assim:

$$E(X^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \exp\Big\{ -\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big\} dx,$$

utilizando novamente $u = \frac{x-\mu}{\sigma}$, organizando a expressão, obtemos:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2\sigma^2 \exp\Big\{-\frac{u^2}{2}\Big\} du + \mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{u^2}{2}\Big\} du + 2\sigma\mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \exp\Big\{-\frac{u^2}{2}\Big\} du. \tag{2.23}$$

Note que a terceira integral é uma função ímpar, então o resultado de sua integração é nulo. Para segunda integral vemos que $u \sim$ Normal $(0, 1)$, desta forma o resultado da operação é um. Resta apenas a primeira integral que será resolvida com integração por partes.

Fazendo $u = u$ e $dv = u\exp(-\frac{u}{2})$, obtemos logo $v = -\exp -\frac{u^2}{2}$. Aplicando o método de integração por partes e substituindo na equação 2.23, temos que $E(X^2) = \sigma^2 + \mu^2$, e consequentemente:

$$VAR(X) = \sigma^2. \tag{2.24}$$

2.5.1.2 Estimação de Verossimilhança para Distribuição Normal

Fazendo uso da Definição 1, temos que a função de verossimilhança para a distribuição Normal é apresentada por:

$$L(\mu, \sigma^2; x) = \Big(\sigma^2 2\pi\Big)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{n}(x_i - \mu)^2\right\}. \tag{2.25}$$

Linearizando, derivando e igualando a zero equação 2.25, têm-se os parâmetros de verrosimilhança para distribuição normal, que são:

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}, \\ \hat{\sigma^2} &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}. \end{aligned} \tag{2.26}$$

Vale ressaltar que diferente das distribuições que veremos, o MV para a distribuição normal resulta em estimativas diretas, não sendo necessário o uso de métodos iterativos para encontrar seus valores.

2.5.2 Distribuição Gama

A distribuição Gama é amplamente utilizada para modelar dados em diversas áreas, como por exemplo, descrever o tempo de vida útil de um equipamento elétrico, o ciclo de precipitação, e também um processo de fabricação de certo material.

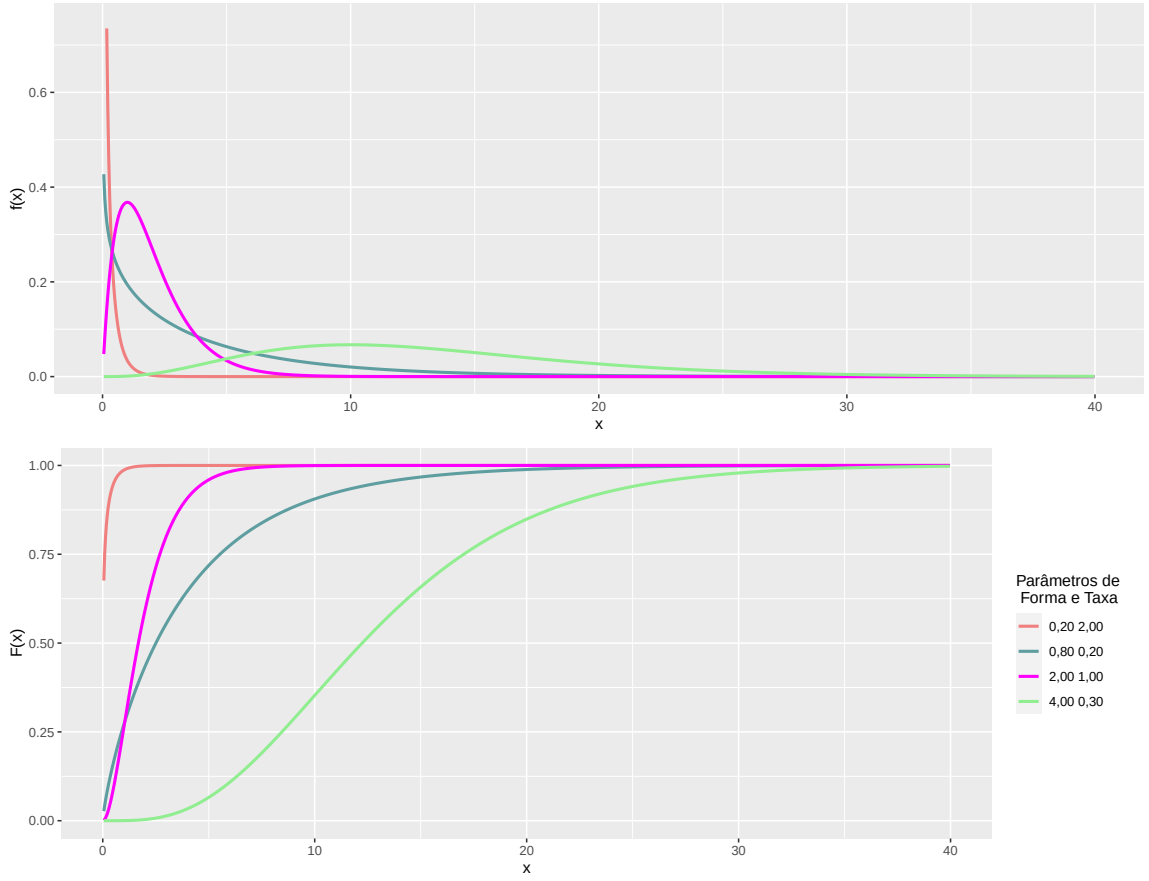
Assim, uma variável aleatória contínua X tem distribuição de probabilidade Gama com parâmetros $\phi > 0$ e $\lambda > 0$, $X \sim Gama(\phi, \lambda)$ se sua função densidade de probabilidade e, consequentemente, sua função de distribuição acumulada são definidas como se segue:

$$f(x|\phi,\lambda) = \frac{\lambda^{\phi} x^{\phi-1} e^{-\lambda x}}{\Gamma(\phi)} \quad e \quad F(x|\phi,\lambda) = \frac{\Gamma(\phi,\lambda x)}{\Gamma(\phi)}, \tag{2.27}$$

com $x > 0$. Sendo $\phi > 0$ o parâmetro de forma, $\lambda > 0$ o parâmetro de taxa, $\Gamma(\phi) = \int_x^{\infty} t^{\phi-1} e^{-t} dt$, a função Gama, e $\Gamma(\phi, x\lambda) = \int_0^{x\lambda} t^{\phi-1} e^{-t} dt$ a função gama incompleta.

Os gráficos para diferentes pares de parâmetros da FDP e FDA da distribuição Gama são apresentados na Figura 2.2.

Figura 2.2: Função de distribuição de probabilidade (A) e Função densidade Acumulada (B) da distribuição Gama.



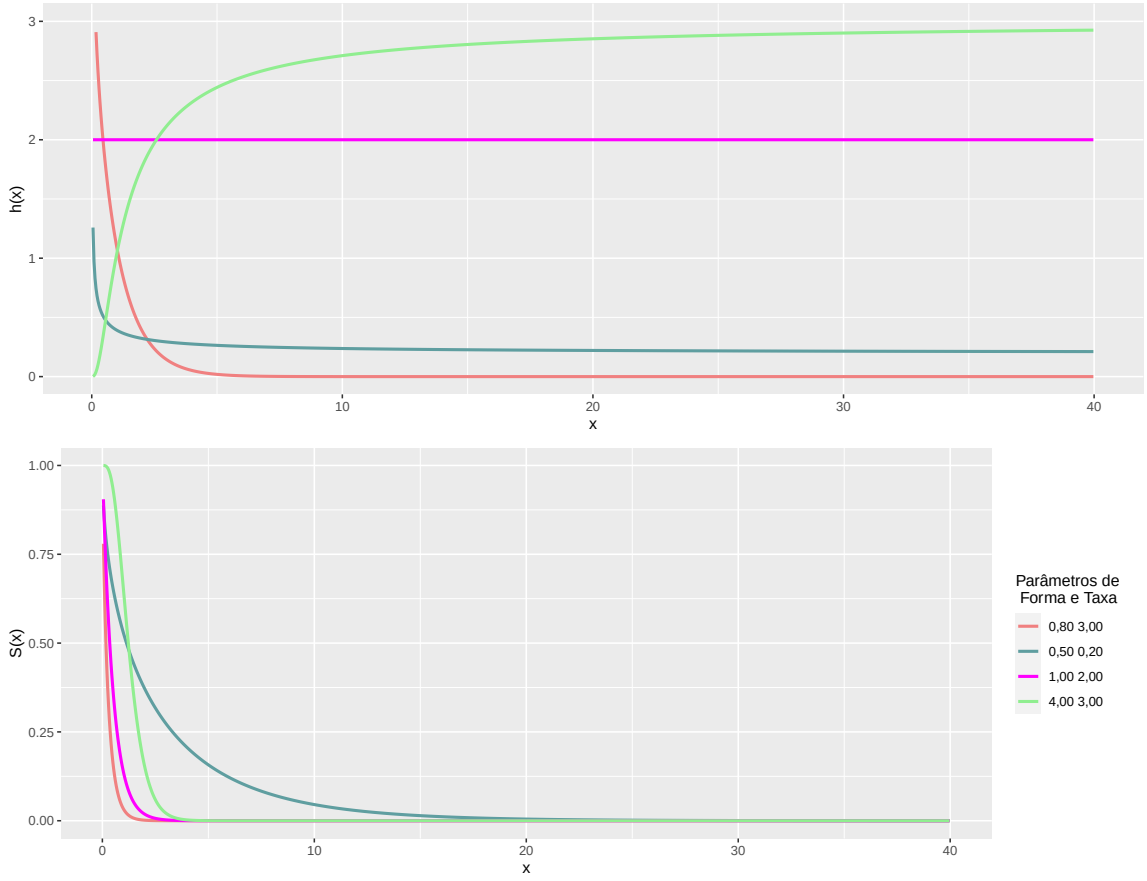
Suas funções de confiabilidade e risco são dadas, respectivamente, por:

$$
\begin{aligned}
R(x|\phi,\lambda) &= 1 - \frac{\Gamma(\phi,\lambda x)}{\Gamma(\phi)}, \\
h(x|\phi,\lambda) &= \frac{\lambda^{\phi} x^{\phi-1}\exp(-\lambda x)}{\Gamma(\phi,\lambda x)\Gamma(\phi)},
\end{aligned}
\tag{2.28}
$$

com x, ϕ, $\lambda > 0$ e $\Gamma(\phi,\lambda x)$ a função gama incompleta.

Da mesma forma, a seguir têm-se as formas das funções de sobrevivência e risco para diferentes pares de parâmetros da distribuição Gama.

Figura 2.3: Função de Risco (A) e Função de Sobrevivência (B) da distribuição Gama.



Observe que para $\phi > 1$ o risco, cresce de forma monótona, para $0 < \phi < 1$ ocorre o contrário, o risco decresce monotonicamente e por fim para $\phi = 1$ tem-se o risco constante, o que era esperado uma vez que $Gama(1,\lambda) = Exponencial(\lambda)$.

2.5.2.1 Medidas da Distribuição Gama

Algumas medidas da distribuição Gama, geradas a partir da função geratriz de momentos que é definida por:

$$
\begin{aligned}
M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx}\frac{\lambda^{\phi}x^{\phi-1}e^{-\lambda x}}{\Gamma(\phi)}dx = \\
\frac{\lambda^{\phi}}{\Gamma(\phi)}\int_0^{\infty} x^{\phi-1}e^{-x(\lambda-t)}dx = \left(\frac{\lambda}{\lambda-t}\right)^{\phi}.
\end{aligned}
\tag{2.29}
$$

Para encontramos a esperança e variância da distribuição em questão, vamos derivar a função geratriz nos primeiro e segundo momentos, temos:

$$M_X^{'}(t) = \phi(\lambda - t)^{-(\phi+1)}\lambda^{\phi} = \frac{\phi}{\lambda - t}\left(\frac{\lambda}{\lambda - t}\right)^{\phi},$$
$$M_X^{''}(t) = \frac{\phi(1+\phi)}{(\lambda - t)^2}\left(\frac{\lambda}{\lambda - t}\right)^{\phi}. \tag{2.30}$$

Obtemos que:

$$E(X) = M_X^{'}(0) = \frac{\phi}{\lambda},$$

e sabemos também que $Var(X) = E(X^2) - (E(X))^2$, assim:

$$Var(X) = M_X^{''}(0) - (M_X^{'}(0))^2 = \frac{\phi}{\lambda^2}.$$

2.5.2.2 Estimação de Verossimilhança para Distribuição Gama

Usando a Definição 1, a função de verossimilhança para a distribuição Gama é dada por:

$$L(\phi, \lambda; \mathbf{x}) = \left(\frac{\lambda^{\phi}}{\mathbf{\Gamma}(\phi)}\right)^{\mathbf{n}} \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i^{\phi-1}} \exp\left(-\lambda \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i}\right). \tag{2.31}$$

As derivadas parciais de cada parâmetros igualadas à zero, resultam neste sistema não-linear :

$$\hat{\lambda} = \frac{n\hat{\phi}}{\sum_{i=1}^{n} x_i},$$
$$\psi(\hat{\phi}) = \log(\hat{\lambda}) - \log(\bar{x}), \tag{2.32}$$

com $\psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$, a função digamma. Para encontramos os estimadores precisaremos de métodos iterativos, uma vez que esta é uma equação não-linear.

Métodos iterativos, como o exemplo apresentado na subseção 2.3.2, possuem um alto valor computacional, baseados nesta dor no capítulo 3 é exposto um método de estimação alternativo para encontrar a estimativa dos parâmetros da distribuição Gama com uma forma fechada, ou seja, de uma forma direta.

2.5.3 Distribuição Weibull

Proposta por Weibull em 1951 a distribuição é amplamente usada em análise de confiabilidade e utilizada em diversas aplicações na área industrial. Seja T uma variável aleatória contínua, ela possui distribuição Weibull se apresentar as seguintes funções de distribuição acumulada e densidade de probabilidade:

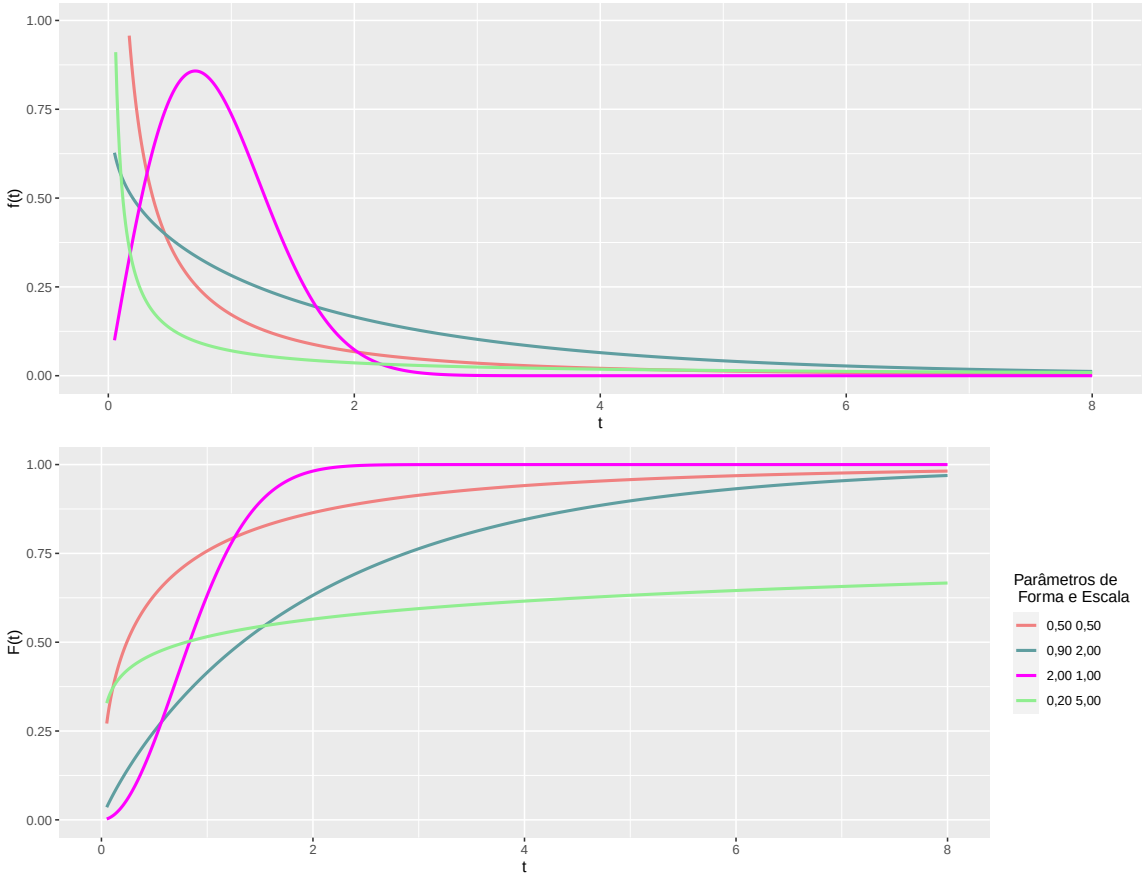
$$F(t|\lambda, \delta) = 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{\delta}\right] \quad e \quad f(t|\lambda, \delta) = \frac{\delta}{\lambda^{\delta}} t^{\delta-1} \exp\left[-\left(\frac{t}{\lambda}\right)^{\delta}\right], \tag{2.33}$$

para $t > 0$, sendo $\delta > 0$ o parâmetro de forma, e $\lambda > 0$ o de escala.

Podemos destacar que a distribuição Weibull é caso particular de algumas distribuições, como por exemplo, a distribuição Exponencial, quando $\delta = 1$ e também a distribuição de Rayleigh quando $\delta = 2$.

Na Figura 2.4, vemos um gráfico das distribuições acumulada e densidade de probabilidade para diferentes pares de parâmetros.

Figura 2.4: Função de distribuição de probabilidade (A) e Função densidade acumulada (B) da distribuição Weibull.



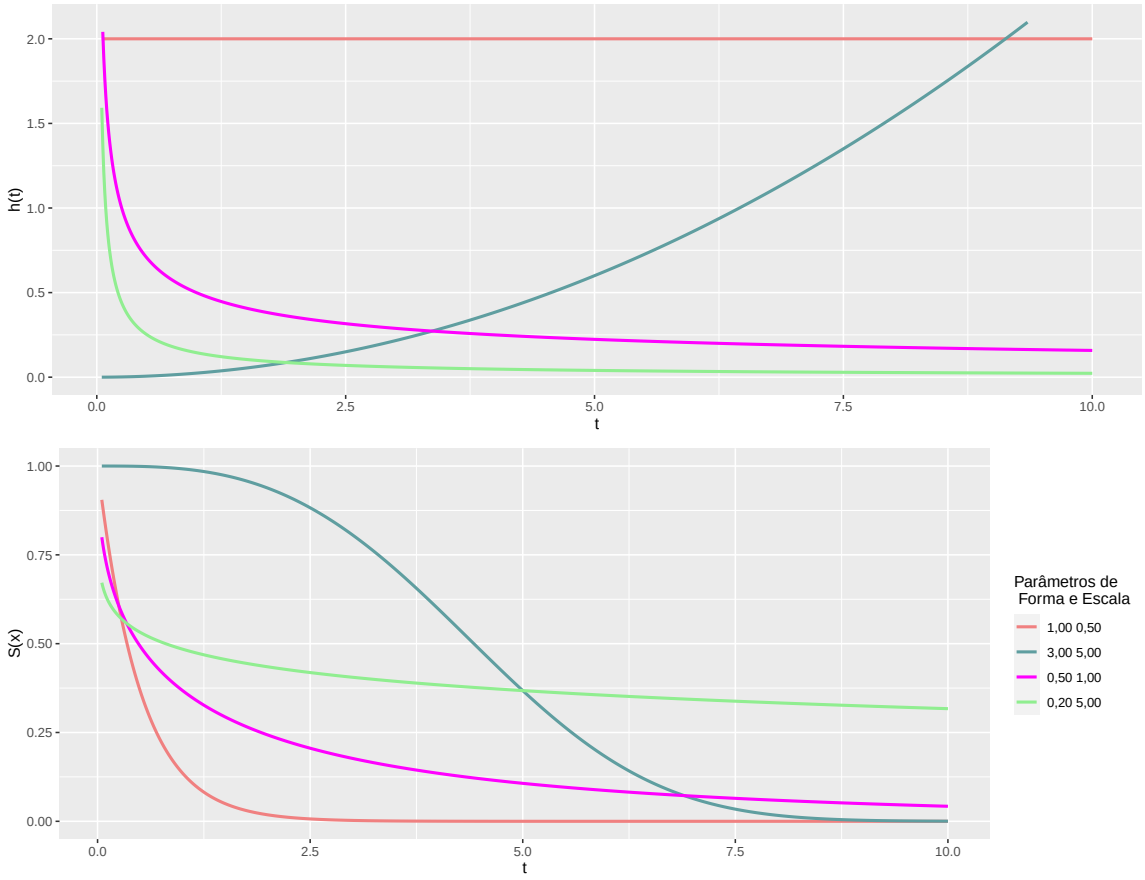
Da mesma forma, apresenta-se agora as funções de confiabilidade e de risco para a distribuição:

$$R(t|\lambda,\delta) = \exp\left[-\left(\frac{t}{\lambda}\right)^{\delta}\right] \quad \text{e} \quad h(t|\lambda,\delta) = \frac{\delta}{\lambda^{\delta}}t^{\delta-1}, \tag{2.34}$$

sendo $t > 0$ e λ, $\delta > 0$.

A Figura 2.5 traz algumas formas da função de sobrevivência e de risco para a distribuição de Weibull. Nota-se que para $\delta < 1$ o risco é decrescente, para $\delta > 1$ crescente, e por fim para $\delta = 1$ temos o risco constante.

Figura 2.5: Função de Risco (A) e Função de sobrevivência (B) da distribuição Weibull.



2.5.3.1 Medidas da Distribuição Weibull

Diferente da distribuição Gama, a função geratriz de momentos para a distribuição Weibull não possui forma fechada. Deste modo, encontramos o k-ésimo momento da seguinte forma:

$$E(T^k) = \int_0^\infty t^k \frac{\delta}{\lambda^\delta} t^{\delta-1} exp\left[-\left(\frac{t}{\lambda}\right)^\delta \right] dt, \tag{2.35}$$

vamos usar a seguinte substituição $u = (\frac{t}{\lambda})^\delta$, assim obtemos:

$$du = \frac{\delta}{\lambda^\delta} t^{\delta-1} dt, \quad t = u^{1/\delta}\lambda.$$

Retornando a equação (2.35), e fazendo os devidos ajustes, tem-se:

$$E(T^k) = \lambda^k \int_0^\infty u^{k/\delta} e^{-u} du = \lambda^k \Gamma\left(\frac{k}{\delta} + 1\right). \tag{2.36}$$

E finalmente, a esperança e variância da distribuição são:

$$E(T) = \lambda\Gamma\left(\frac{1}{\delta} + 1\right),$$
$$V(T) = \lambda^2 \left[\Gamma\left(\frac{2}{\delta} + 1\right) - \Gamma^2\left(\frac{1}{\delta} + 1\right) \right].$$

2.5.3.2 Estimação de Verossimilhança para Distribuição Weibull

A função de verossimilhança para a distribuição Weibull, utilizando a Definição 1, é dada por:

$$L(\lambda, \delta; \boldsymbol{t}) = \left(\frac{\delta}{\lambda^{\delta}}\right)^{n} \prod_{i=1}^{n} t^{\delta-1} \exp\left[-\left(\frac{\sum_{i=1}^{n} t}{\lambda}\right)^{\delta}\right]. \tag{2.37}$$

Realizando todo o processo descrito na seção 2.3, leva ao sistema de equações não-lineares dado a seguir:

$$\begin{cases} \frac{\hat{\delta}}{\hat{\lambda}}\left(\frac{\sum_{i=1}^{n} t_i^{\hat{\delta}}}{\hat{\lambda}^{\hat{\delta}}} - n\right) = 0; \\ n(\hat{\delta}^{-1} - \log\hat{\lambda}) + \sum_{i=1}^{n} \log(t_i) - \hat{\delta}\sum_{i=1}^{n}\left(\frac{t_i}{\hat{\lambda}}\right)^{\hat{\delta}-1} = 0. \end{cases} \tag{2.38}$$

A solução do sistema, que é obtida através de métodos numéricos, resulta nos estimadores de máxima verossimilhança para a distribuição Weibull, porém como dito no caso da distribuição Gama, requer um custo computacional. Desta forma, no capítulo 4 será apresentado um método de estimação direto, que estime os parâmetros para a distribuição em questão.

Capítulo

3

Método de Verossimilhança Modificado para Distribuição Gama

Para algumas distribuições, como é o caso da distribuição Gama, o método de estimação de verossimilhança não possui forma fechada para a estimativa dos parâmetros, sendo necessário o uso de métodos iterativos, como o de Newton-Raphson, para encontrá-los.

Computacionalmente falando, usar recursos iterativos é custoso, pensando nisto Ramos *et al.* [17] propuseram o método de verossimilhança modificado que resulta em estimativas fortemente consistentes e com distribuição normal assintótica. A motivação é obter expressões em forma fechada para se estimar os parâmetros das distribuições com a função de densidade $f(x|\boldsymbol{\theta})$.

Para formular o método, os autores definiram o vetor paramétrico $\Theta \subset \mathbb{R}^s$, um conjunto aberto, que contêm os valores $\boldsymbol{\theta}$ que se deseja estimar, e $\aleph \subset \mathbb{R}^r, 0 \leq r \leq s$, outro conjunto aberto que contém o parâmetro adicional α, que será utilizado para obter os estimadores de forma direta. E ainda, o conjunto Ω representando o espaço amostral e $\chi_\alpha \subset \mathbb{R}$ um conjunto mensurável que depende do paramêtro adicional α que representa o espaço dos dados denotado por $\chi = \chi_{\alpha_0}$.

Dado o parâmetro $\boldsymbol{\theta_0} \in \Theta$ que será estimado, fixando um $\alpha_0 \in \aleph$, vamos supor que $X_1, X_2, ..., X_n$ são variáveis aleatórias *iid*, com uma função de densidade estritamente positiva dada $f(x|\boldsymbol{\theta_0})$ definida para todos os $x \in \chi$. Considere que a função $f(x|\boldsymbol{\theta})$ possa ser vista como um caso especial de outra função de densidade com parâmetros adicionais, $g(x|\boldsymbol{\theta}, \boldsymbol{\alpha})$ também definida para todos $x \in \chi_\alpha$, ou seja, $f(x|\boldsymbol{\theta}) = g(x|\boldsymbol{\theta}, \boldsymbol{\alpha_0})$ para todos os $x \in \chi$ e $\boldsymbol{\theta} \in \Theta$. Então, define-se as equações de máxima verossimilhança modificada para g com $\alpha = \alpha_0$ como:

$$\begin{aligned} &\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log\ g(x_i|\boldsymbol{\theta}, \boldsymbol{\alpha}_0), \qquad 0 \leq j \leq s - r - 1, \\ &\sum_{i=1}^{n} \frac{\partial}{\partial \alpha_j} \log\ g(x_i|\boldsymbol{\theta}, \boldsymbol{\alpha}_0), \qquad 0 \leq j \leq r - 1. \end{aligned} \tag{3.1}$$

O Método MVM gera estimadores: invariantes sob transformações um-a-um, consistentes. E ainda, tem distribuição normal assintótica. Para mais detalhes consultar Ramos *et al.* [17].

O método de verossimilhança modificado para a distribuição Gama é aplicado em sua distribuição generalizada, a distribuição Gama Generalizada, cuja função densidade de probabilidade é dada por:

$$g(x|\lambda,\phi,\alpha) = \frac{\alpha}{\Gamma(\phi)}\Big(\frac{\phi}{\lambda}\Big)^{\phi} x^{\alpha\phi-1}\exp\Big(-\frac{\phi}{\lambda}x^{\alpha}\Big), \tag{3.2}$$

para $x > 0$, sendo o parâmetro $\lambda > 0$ de escala, e $\phi > 0$ e $\alpha > 0$ parâmetros de forma.

Sendo assim, considerando $\alpha = 1$, uma vez que $f(x|\lambda,\phi) = g(x|\lambda,\phi,\alpha = 1)$ e ainda para $n \geq 2$ $s = r = 1$, aplicando (3.1) temos que as equações do método de máxima verrosimilhança modificado para a distribuição Gama são dadas por:

$$\begin{aligned}
\sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\log\ g(x_i|\lambda,\phi,\alpha_0) =& \frac{\phi}{\lambda^2}\sum_{i=1}^{n}x_i - \frac{n\phi}{\lambda},\\
\sum_{i=1}^{n}\frac{\partial}{\partial\alpha_j}\log\ g(x_i|\lambda,\phi,\alpha_0) =& \phi\left(\sum_{i=1}^{n}\log(x_i) - \frac{\sum_{i=1}^{n}x_i\log(x_i)}{\lambda}\right) + n.
\end{aligned} \tag{3.3}$$

Prosseguindo com o método, igualando as equações em (3.3) à zero, e seguindo Louzada *et.al* [12], considerando que a igualdade $X_1 = X_2 = ... = X_n$ não ocorre, então $n\sum_{i=1}^{n}x_i log(x_i) - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}log(x_i) \neq 0$. E por fim, os estimadores de MVM para a distribuição Gama são:

$$\hat{\lambda}_{MVM} = \frac{\sum_{i=1}^{n}x_i}{n} \quad \text{e} \quad \hat{\phi}_{MVM} = \frac{n\sum_{i=1}^{n}x_i}{n\sum_{i=1}^{n}x_i log(x_i) - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}log(x_i)}. \tag{3.4}$$

E ainda, para n suficientemente grande, temos que :

$$\begin{aligned}
\sqrt{n}(\hat{\lambda}-\lambda) &\sim^{n\to\infty} N\Big[0, \frac{\lambda^2}{\phi}\Big]\\
\sqrt{n}(\hat{\phi}-\phi) &\sim^{n\to\infty} N\Big[0, \phi^3\psi'(\phi+1)+\phi^2\Big],
\end{aligned} \tag{3.5}$$

com $\psi'(\phi+1) = \frac{\Gamma'(\phi+1)}{\Gamma(\phi+1)}$. Ou seja, as estimativas são não viciadas, consistentes e assintoticamente normais.

Capítulo

4

Método de Estimação L - Momentos para Distribuição Weibull

O método de estimação de verossimilhança para a Distribuição Weibull não possui uma forma fechada para a estimação de parâmetros sendo necessário o uso de métodos numéricos para estimá-los. Desta forma, Hosking [7] sugeriu o método de estimação L-Momentos (MLM), que em suas aplicações se mostrou robusto e eficiente como verifica-se em Teimouri *et al.* [23].

O método foi desenvolvido a partir de estatísticas de ordem e momentos amostrais. Assim, criou-se funções lineares empíricas, que sofrem menos com os efeitos da variabilidade presente na amostra, levando à uma maior robustez na presença de outliers, oferecendo inferências seguras mesmo para amostras pequenas, como pode-se ver em Elsayed e Allan [5].

Dado T uma variável aleatória real com função de distribuição acumulada e quantil dadas por $F(t)$ e $t(F)$, e ainda $T_{1:n} \leq T_{2:n} \leq ... \leq T_{n:n}$ as estatísticas de ordem para uma amostra aleatória de tamanho n da distribuição T , os L- Momentos são definidos como:

$$\lambda^r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(T_{r-k:r}), \qquad \text{para} \quad r = 1, 2, ..., n. \tag{4.1}$$

Vale ressaltar que os L-momentos são casos especiais dos momentos ponderados por probabilidade, assim como os momentos ordinários, apresentados por Greenwood *et al.* [6]. Eles são dados por $M_{p,r,s} = E[T^p(F(T))^r(1-F(T))^s]$, note que $M_{1,0,s}$ e $M_{1,r,0}$ podem ser vista como combinações lineares de L-momentos.

Hosking [7] se utiliza das seguintes justificativas para utilizar os L-momentos com o objetivo de descrever distribuições de probabilidade, que são:

- Os L-momentos λ_r , $r = 1, 2, ..., n$ de uma variável aleatória de valor real T existem, se e somente se, T tiver média finita;
- Uma distribuição cuja média existe é caracterizada por seus L-momentos (λ_r), $r = 1, 2, ..., n$.

Para mais detalhes, verifique Hosking [7].

Desta forma, temos que o r-ésimo L-Momento amostral é dado por:

$$l_r = r^{-1}\binom{n}{r}^{-1}\sum_{i=1}^{n} t_{i:n}\sum_{k=0}^{r-1}(-1)^k\binom{(r-1)}{k}\binom{(n-i)}{k}\binom{i-1}{r-k-1}, \qquad r = 1, 2, ..., n. \tag{4.2}$$

Para a distribuição Weibull, com função densidade de probabilidade dada em (2.33) aplicando a definição de L-momento populacional e amostral apresentados nas equações (4.1) e (4.2) temos que :

$$\begin{aligned} \lambda_1 &= \lambda\Gamma\Big(\frac{1}{\delta}+1\Big), \qquad & \lambda_2 &= \lambda\Gamma\Big(\frac{1}{\delta}+1\Big)\Big[1-\frac{1}{2^{-\delta}}\Big] \\ l_1 &= \sum_{i=1}^{n}\frac{t_{i:n}}{n} = \bar{t}, \quad & l_2 &= \frac{2}{n(n-1)}\sum_{i=1}^{n}(i-1)t_{i:n} - \bar{t}. \end{aligned} \tag{4.3}$$

Igualando os momentos populacionais aos amostrais, obtemos as estimativas L-Momentos para a distribuição Weibull, apresentadas a seguir:

$$\begin{aligned} \hat{\delta}_{LM} &= -\frac{log(2)}{log\Big(1-\frac{l_2}{l_1}\Big)} \\ \hat{\lambda}_{LM} &= \frac{l_1}{\Gamma\Big(\frac{1}{\hat{\delta}}+1\Big)} \end{aligned} \tag{4.4}$$

Observe que os parâmetros são encontrados de forma direta, não sendo necessário o uso de métodos iterativos como ocorre com método de máxima verossimilhança para esta distribuição, diminuindo o custo e tempo computacional.

Capítulo 5

Proposta Gráfica

Este capítulo é reservado para expormos a proposta gráfica. O objetivo apresentado neste trabalho é verificar como em um processo, em tempo real, o índice C_{pk} verifica oscilações, utilizando os metódos de estimações direta.

Para isto, criamos uma função no software livre R, que chamamos de *cpk.plot(data*, n, m, LSE, LIE, *dist)* que recebe os dados do usuário, calcula os valores de $C_{pk}s$ e também seus respectivos intervalos de 95% de confiança para as m amostras, utilizando mil reamostragens de bootstrap, sendo seu *output* um gráfico com faixas de atenção que indica se o processo está descontrolado, em ponto de atenção ou se o processo é capaz de atentar as especificações, representados pelas cores vermelha, cinza e azul, baseados no valor de C_{pk} calculados no decorrer do processo.

O *script* da função é apresentada no Apêndice B. Seus argumentos são :

- *data*− são os observações do processo;
- m− quantidade de amostras;
- n− elementos em cada amostra;
- LSE− limite superior de especificação;
- LIE− limite inferior de especificação;
- $dist = (\text{"}Gamma\text{"}, \text{"}Weibull\text{"}, \text{"}Normal\text{"})$− distribuições presentes na função. Uma deve ser escolhida pelo usuário.

As faixas de atenção, que definem a situação do processo no decorrer do tempo, foram delimitadas pelos intervalos abaixo, baseados na interpretação do índice dado por Louzada *et al* [13]:

- Área vermelha : para índices $C_{pk} < 1$, indicando um processo não capaz de atender as especificações;
- Área cinza: com $1,00 \leq C_{pk} < 1,33$, apontando um processo parcialmente capaz em atender as normas;
- Área azul : para valores de $C_{pk} \geq 1,33$, demonstrando um processo inteiramente capaz de atender as especificações dos órgãos reguladores.

Acionando a função usando *cpk.plot(data = x, n = 5, m = 20, LIE = 0, LSE = 12, dist = "Gamma")*, temos um processo, armazenado no vetor x, que segue a distribuição Gama, com parâmetro de forma e taxa iguais a dois, no qual há 20 amostras, cada uma com tamanhos iguais a 5, além de que os limites de especificação superior e inferior são 12 e 0, respectivamente, que resulta nas Figura 5.1 e Tabela 5.1.

Figura 5.1: Saída Gráfica da Função *cpk.plot()*

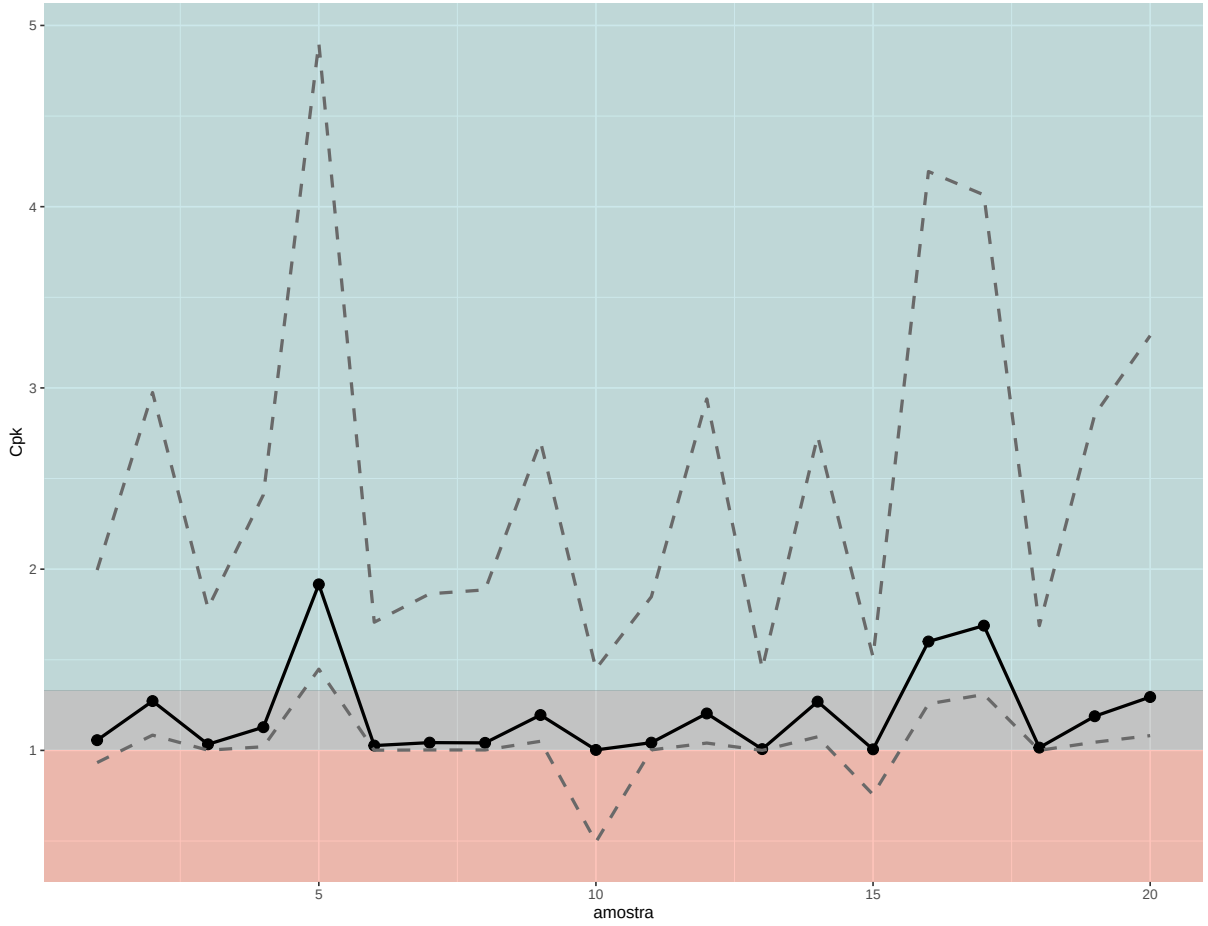


Tabela 5.1: Saída Numérica da Função *cpk.plot()*

	I.C Inferior	Cpk - MVM	I.C Superior
1	0.93	1.06	2.00
2	1.08	1.27	2.97
3	1.00	1.03	1.79
4	1.02	1.13	2.41
5	1.45	1.92	4.90
6	1.00	1.03	1.71
7	1.00	1.04	1.86
8	1.00	1.04	1.89
9	1.05	1.20	2.70
10	0.49	1.00	1.45
11	1.00	1.04	1.85
12	1.04	1.20	2.94
13	1.00	1.01	1.45
14	1.07	1.27	2.73
15	0.76	1.01	1.52
16	1.26	1.60	4.19
17	1.31	1.69	4.07
18	1.00	1.02	1.69
19	1.05	1.19	2.85
20	1.08	1.29	3.29

Observe que em todas as amostras, o processo é considerado parcialmente em capaz atender as normas reguladores. O mesmo ocorre com seu valores intervalares. O resultado apresentado era esperado uma vez que todo este processo possui $C_{pk} = 1.03$, um valor que indica controle porém com cautela, uma vez que está quase na fronteira entre o capaz e incapaz.

O *script* da função é apresentada no Apêndice B, para possíveis aplicações.

Capítulo

6

Simulação

Este capítulo é dedicado ao estudo de simulação. O mesmo foi realizado com o objetivo de verificar como o índice C_{pk} captura as mudanças durante um processo em tempo real utilizando a estimação direta.

Sendo assim, simulou-se processos que seguem distribuições Normal, Gama e Weibull, com três cenários possíveis em cada. O primeiro dito cenário fronteira, o segundo cenário ótimo e por fim, o cenário descontrole. Cada cenário possui um valor para o índice C_{pk} que será explicitado adiante.

De forma geral, o cenário ótimo apresenta um processo com o índice C_{pk} próximo a 1,33 , para o cenário descontrolado o índice é abaixo de 1,00, e o último, fronteira a métrica fica próximo a 1,00.

Desta forma, as simulações foram realizadas com 10000 replicações de subamostras com tamanhos $n = (5, 10, 15, 20, 25, 30, 35, 40, 45, 50)$, nas quais o valor do índice foi calculado e identificado se aquela subamostra, de uma amostra de tamanho $m = n * 100$, era ou não capaz de atender as especificações.

Para os cenários chamados de fronteira e ótimo em parte do processo causamos um acréscimo de cinco unidades, com o objetivo de descontrolar aquela parte, e esperar que o índice aponte a mudança. Já para o cenário descontrole, não foi necessário causar a pertubação, uma vez que, todo o processo já era incapaz.

Apresentaremos inicialmente a eficácia do índice em identificar a mudança para valores pontuais de C_{pk} e, após será apresentado para valores intervalares. Ambos para os três cenários possíveis.

Por fim, as estimações realizadas foram pelos métodos diretos: MV para a distribuição Normal, o MVM para a Gama, e MLM para Weibull, e também o método de verossimilhança para as duas últimas distribuições como um método indireto de estimação, uma vez que, para elas o método não apresenta soluções explícitas.

6.1 Estudo de Simulação classificação de valores pontuais de C_{pk}

Esta seção é dividida em três partes, a primeira sendo para dados que seguem a distribuição Gaussiana, a segunda reservada para os que possuem distribuição Gama e a última para os que são distribuídos conforme a Weibull.

Em todas para medir a classificação do modelo utilizamos medidas de acurácia, valor predito positivo (precision), sensibilidade (recall) e por fim, especificidade. Para definição

das medidas citadas anteriormente considere na Tabela 6.1 a matriz de confusão para o modelo:

Tabela 6.1: Matriz de Confusão para o modelo

Predito Pelo Modelo	Mundo Real	
	Índice Descontrolado	Índice Controlado
Índice Descontrolado	VP	FP
Índice Controlado	FN	VN

Desta forma, o valor verdadeiro positivo (VP), indica que o modelo identificou a amostra descontrolada corretamente baseado no valor do índice C_{pk}. Assim as medidas de classificação são dadas por:

- Acurácia: proporção de acertos, ou seja, amostras identificados corretamente como descontroladas pelo modelo de classificação. Representada na tabela por AC dada por:
$$AC = \frac{VN + VP}{VN + FP + VP + FN};$$
- Valor Predito Positivo (VPP): uma medida da proporção de índices descontrolados classificado segundo o modelo. Essa medida é dada por
$$VPP = \frac{VP}{VP + FP};$$
- Sensibilidade(SB): que representa a proporção de índices descontrolados classificados corretamente:
$$SB = \frac{VP}{VP + FN};$$
- Especificidade (EP): a proporção de índices controlados classificados corretamente, representado na tabela por EP, e é dado por:
$$EP = \frac{VN}{VN + VP}.$$

6.1.1 Distribuição Normal

Os cenários para a distribuição Normal foram gerados como segue-se: as amostras do cenário fronteira foram geradas com média 35 e variância 2 já os limites de especificação iguais à 30,71 e 50 o que retornou um valor de $C_{pk} = 1,011163$. Para o cenário ótimo, o valor do índice $C_{pk} = 1,33$ com limites 10 e 18,99, $\mu = 15$ e $\sigma = 1$. E para o descontrole, os parâmetros para a geração da amostra foram $\mu = 1, \sigma = 0,5$, com limites de especificação 2 e 0,05, o que resultou um $C_{pk} = 0,6333$.

Lembrando que como os dados seguem a distribuição normal, o C_{pk} usado aqui é o apresentado na subseção 2.2.1 na equação 2.9.

As medidas de classificação pontual para a distribuição normal são apresentadas na Tabela 6.2.

Tabela 6.2: Estudo de simulação de classificação para valores pontuais de $C'_{pk}s$ para distribuição Normal.

Cenário	n	MV			
		AC	VPP	SB	EP
Fronteira	5	0,4027	0,4461	0,8053	0,0000
	10	0,4072	0,4489	0,8145	0,0000
	15	0,3970	0,4426	0,7940	0,0000
	20	0,3956	0,4417	0,7912	0,0000
	25	0,3875	0,4366	0,7750	0,0000
	30	0,3776	0,4303	0,7552	0,0000
	35	0,3671	0,4234	0,7342	0,0000
	40	0,3673	0,4235	0,7346	0,0000
	45	0,3595	0,4183	0,7190	0,0000
	50	0,3555	0,4155	0,7110	0,0000
Ótimo	5	0,5003	0,5001	1,0000	0,0005
	10	0,5085	0,5042	1,0000	0,0166
	15	0,5442	0,5231	1,0000	0,0811
	20	0,6146	0,5647	1,0000	0,1864
	25	0,7008	0,6256	1,0000	0,2865
	30	0,7824	0,6967	1,0000	0,3609
	35	0,8457	0,7642	1,0000	0,4088
	40	0,9006	0,8341	1,0000	0,4448
	45	0,9344	0,8839	1,0000	0,4649
	50	0,9558	0,9187	1,0000	0,4769
Descontrolado	5	0,5000	0,5000	1,0000	-
	10	0,5000	0,5000	1,0000	-
	15	0,5000	0,5000	1,0000	-
	20	0,5000	0,5000	1,0000	-
	25	0,5000	0,5000	1,0000	-
	30	0,5000	0,5000	1,0000	-
	35	0,5000	0,5000	1,0000	-
	40	0,5000	0,5000	1,0000	-
	45	0,5000	0,5000	1,0000	-
	50	0,5000	0,5000	1,0000	-

Note que, para o cenário descontrolado independente do tamanho da subamostra o modelo acertou em torno de 50%. Por outro lado, não classificou nenhum verdadeiro negativo, uma vez que não há, e por isto a especificidade foi representada por um "-". Para os outros dois cenários podemos identificar que ao aumentar o tamanho da subamostra a acurácia também aumentava.

Particularmente para o cenário fronteira, a sensibilidade, que é a medida responsável por quantificar os itens descontrolados classificados corretamente, foi maior para tamanhos de subamostras menores. E ainda, para este mesmo cenário não teve índices controlados classificados corretamente, olhando a matriz de confusão identificamos que não houve, em nenhuma iteração, itens ditos controlados, lembrando que estamos estimando um valor de índice entre o que é considerado capaz e não capaz.

Para o cenário ótimo, destacamos que todos os itens descontrolados, que é o nosso foco, foram identificados corretamente, o mesmo ocorre no cenário descontrolado.

6.1.2 Distribuição Gama

Iniciamos com a apresentação dos cenários. O cenário fronteira, com $C^*_{pk} = 1,014078$, e parâmetros e limites de especificação dados por $\lambda = 2$, $\phi = 1$, $LIE = 0,03$ e $LSE = 12,5$. Por sua vez, o cenário ótimo com $C^*_{pk} = 1,33$, parâmetros $\lambda = 50$, $\phi = 5$ e limites de especificação inferior e superior iguais a 0 e 16,37847. E finalmente, o cenário descontrolado $\lambda = 55, \phi = 10, LIE = 4$ e $LSE = 16,37847$.

As medidas de classificação pontuais são apresentadas na Tabela 6.3.

Tabela 6.3: Estudo de simulação de classificação para valores pontuais de C^*_{pk} para distribuição Gama.

Cenário	n	MVM				MV			
		AC	VPP	SB	EP	AC	VPP	SB	EP
Fronteira	5	0,4642	0,4815	0,9285	0,0000	0,4996	0,4998	0,9910	0,0000
	10	0,4723	0,4858	0,9446	0,0000	0,4995	0,4997	0,9990	0,0000
	15	0,4597	0,4790	0,9182	0,0013	0,4993	0,4996	0,9985	0,0000
	20	0,4377	0,4667	0,8713	0,0048	0,4991	0,4995	0,9982	0,0000
	25	0,4111	0,4505	0,8090	0,0161	0,4988	0,4984	0,9976	0,0000
	30	0,3939	0,4379	0,7489	0,0498	0,4978	0,4989	0,9955	0,0000
	35	0,3713	0,4194	0,6704	0,0971	0,4967	0,4983	0,9934	0,0000
	40	0,3668	0,4099	0,6057	0,1745	0,4961	0,4961	0,9922	0,0000
	45	0,3658	0,3985	0,5271	0,2794	0,4937	0,4968	0,9873	0,0001
	50	0,3686	0,3891	0,4611	0,3746	0,4922	0,4961	0,9842	0,0002
Ótimo	5	0,5001	0,5001	1,0000	0,002	0,5002	0,5001	1,0000	0,0003
	10	0,5014	0,5007	1,0000	0,0028	0,5017	0,5008	1,0000	0,0034
	15	0,5087	0,5004	1,0000	0,0171	0,5088	0,5044	1,0000	0,0173
	20	0,5314	0,5162	1,0000	0,0590	0,5312	0,5161	1,0000	0,0586
	25	0,5742	0,5400	1,0000	0,1291	0,5750	0,5454	1,0000	0,1304
	30	0,6987	0,6240	1,0000	0,2844	0,6363	0,5789	1,0000	0,2115
	35	0,6989	0,6241	1,0000	0,2846	0,6990	0,6241	1,0000	0,2846
	40	0,7628	0,6782	1,0000	0,3445	0,7616	0,6771	1,0000	0,3435
	45	0,8169	0,7320	1,0000	0,3879	0,8169	0,7320	1,0000	0,3879
	50	0,8556	0,7759	1,0000	0,4156	0,8570	0,776	1,0000	0,4166
Descontrolado	5	0,5000	0,5000	1,0000	-	0,5000	0,5000	0,9989	-
	10	0,5000	0,5000	1,0000	-	0,4995	0,4997	1,0000	-
	15	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	20	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	25	0,5000	0,5000	1,0000	-	0,5000	0,5000	0,9999	-
	30	0,5000	0,5000	1,0000	-	0,4995	0,4999	1,0000	-
	35	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	40	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	45	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	50	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-

Para o cenário descontrolado, o valor de especificidade sempre será nulo, uma vez que não há itens controlados nele, e por isto resolvemos representá-lo por " - ".

Para o método de estimação direto, o MVM, note que para o cenário fronteira ($\phi = 2$ e $\lambda = 1$), o modelo não foi capaz de identificar as amostras controladas para as subamostras de tamanhos 5 e 10, um dos motivos pode ser porque o valor de C^*_{pk} gerado é muito próximo de ser considerado não capaz. Por outro lado, o modelo identificou os índices descontrolados corretamente a uma taxa maior que 46%, chegando a passar dos 90% para subamostras menores que quinze. Já o método MV, para a maior parte o modelo não foi capaz de identificar verdadeiros negativos, uma possível justificativa pode ser pelo valor de índice gerado que é $1,014078$, muito próximo de ser considerado não capaz.

Observe que para $\phi = 50$ e $\lambda = 0,2$, o cenário ótimo, para ambos métodos, ao aumentar o tamanho da subamostra melhoram as medidas. Ainda, todos as amostras descontroladas foram identificadas corretamente pelo modelo.

E por fim, o cenário descontrolado, o modelo performou bem, pois não identificou nenhum item controlado, mas sim o contrário, em todas subamostras identificou o descontrole em torno de 100% dos casos, o que era esperado, uma vez q a amostra produz um valor de C_{pk} não capaz. Tal interpretação de aplica para ambos métodos de estimação.

Sendo assim, em busca de melhores resultados, a simulação será feita para valores intervalares do índice. Os resultados serão apresentados na seção 6.2. Vale ressaltar que a métrica utilizada aqui foi o C^*_{pk} apresentado na seção 2.2.2, equação (2.12).

6.1.3 Distribuição Weibull

Os três cenários para a distribuição Weibull foram produzidos da seguinte forma: cenário fronteira com $\alpha = 4, \beta = 2, LIE = 0.369016$ e $LSE = 10$, que resulta um $C_{pk} = 1,01$, o cenário descontrolado $C_{pk} = 0,6943387$, sendo os parâmetros e limites de especificação dados por $\alpha = 2, \beta = 0,5, LIE = 0,14$ e $LSE = 2$, e o cenário ótimo $C_{pk} = 1.33, \alpha = 5, \beta = 3, LIE = 0.1443575$ e $LSE = 5$.

Como na seção anterior, aqui também os resultados pontuais são apresentados primeiro, se encontram na Tabela 6.4. O método de estimação direto para a distribuição Weibull é o método L-Momentos (MLM), apresentado no capítulo 4.

Note que para todos cenários, o comportamento do modelo foi em relação à especifidade, foi parecido, uma vez que, o modelo não foi capaz de identificar os índices controlados classificados corretamente, com exceto para o cenário ótimo, o MLM para amostras maiores que 35 ultrapassou a casa dos 50%, chegando à 92,55% de itens controlados identificados corretamente. Para o cenário descontrolado, era esperado o valor de especificidade nulo, uma vez, que todo o processo é incapaz, para o cenário fronteira, com um olhar pessimista também, lembrando que o C_{pk} gerado é muito próximo de 1.

Independente do modelo de estimação utilizado, a acurácia do modelo para os cenários fronteira e descontrole estão em torno de 50%, já para o cenário restante, ao aumentar a subamostra, o MV performa melhor, porém requer um custo computacional alto.

Para os cenários fronteira e descontrolado, o modelo identificou todos o casos descontrolados. Já para o cenário ótimo ao aumentar o tamanho amostral, utilizando o método MLM, está taxa foi caindo.

Tabela 6.4: Estudo de simulação de classificação para valores pontuais de $C'_{pk}s$ para distribuição Weibull.

Cenário	n	MLM				MV			
		AC	VPP	SB	EP	AC	VPP	SB	EP
Fronteira	5	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	10	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	15	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	20	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	25	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	30	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	35	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	40	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	45	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
	50	0,5000	0,5000	1,0000	0,0000	0,5000	0,5000	1,0000	0,0000
Ótimo	5	0,4994	0,4997	0,9989	0,0000	0,5025	0,5012	1,0000	0,0049
	10	0,4828	0,4913	0,9654	0,0003	0,5220	0,5112	1,0000	0,0421
	15	0.4243	0,4589	0,8447	0,0046	0,5849	0,5464	1,0000	0,1452
	20	0,3556	0,4111	0,6675	0,0614	0,6886	0,6162	1,0000	0,2739
	25	0,3180	0,3640	0,4869	0,2344	0,7855	0,6997	1,0000	0,3634
	30	0,3169	0,3224	0,3325	0,4754	0,8591	0,7801	1,0000	0,4180
	35	0,3518	0.3058	0,2333	0,6684	0,9145	0,8539	1,0000	0,4532
	40	0,3909	0,2897	0,1503	0,8078	0,9471	0,9042	1,0000	0,4720
	45	0,4210	0,2766	0,0978	0,8838	0,9676	0,9391	1,0000	0,4833
	50	0,4512	0,2897	0,0672	0,9255	0,9808	0,9629	1,0000	0,4902
Descontrolado	5	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	10	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	15	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	20	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	25	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	30	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	35	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	40	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	45	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	50	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-

Para o cenário descontrolado, nesta simulação não há índices de $C_{pk} > 1$, ou seja, controlados, por este motivo a especificidade é nula, sendo assim foi escolhido não utilizar esta informação e representá-la por "-" .

Em busca de melhores resultados, realizaremos a simulação para valores intervalares de C_{pk}, apresentadas na Tabela 6.7.

6.2 Simulação para classificação de intervalos de C_{pk}

Nesta seção é apresentado o estudo de simulação para valores intervalares de C_{pk}, uma vez que, para valores pontuais os resultados não foram tão satisfatórios.

Os intervalos para os índices são obtidos a partir da técnica de reamostragem Boostrap apresentada na subseção 2.4 .

Aqui a matriz de confusão é calculada considerando o VP, verdadeiro positivo, como o limite superior do intervalo dado em (2.19) ser menor que 1, ou seja, incapaz. As distribuições e os cenários utilizados continuam os mesmos que os apresentados nos casos pontuais.

6.2.1 Distribuição Normal

Como ocorreu para a simulação pontual, no cenário descontrole nenhum item controlado foi identificado. Para o cenário fronteira, a proporção de acertos foi baixa, uma vez que o modelo não foi capaz de classificar os itens controlados corretamente, uma vez que o cenário está na fronteira entre o capa e incapaz, assim a estimação, aqui, foi subestimada com relação ao C_{pk} gerado, ou seja, abaixo de um.

Tabela 6.5: Estudo de simulação de classificação para intervalos de $C'_{pk}s$ para distribuição Normal.

Cenário	n	MV			
		AC	VPP	SB	EP
Fronteira	5	0,3996	0,4442	0,7991	0,0000
	10	0,4114	0,4514	0,8227	0,0000
	15	0,3999	0,4444	0,7999	0,0000
	20	0,3899	0,4381	0,7998	0,0000
	25	0,3883	0,4371	0,7766	0,0000
	30	0,3780	0,4305	0,7539	0,0000
	35	0,3697	0,4251	0,7393	0,0000
	40	0,3704	0,4256	0,7408	0,0000
	45	0,3567	0,4163	0,7133	0,0000
	50	0,3580	0,4144	0,7076	0,0000
Ótimo	5	0,5007	0,5003	1,0000	0,0013
	10	0,5079	0,5040	1,0000	0,0154
	15	0,5417	0,5218	1,0000	0,0770
	20	0,6166	0,5660	1,0000	0,1890
	25	0,6991	0,6243	1,0000	0,2848
	30	0,7828	0,6971	1,0000	0,3612
	35	0,8486	0,7675	1,0000	0,4108
	40	0,8960	0,8277	1,0000	0,4419
	45	0,9306	0,8781	1,0000	0,4627
	50	0,9569	0,9206	1,0000	0,4775
Descontrolado	5	0,5000	0,5000	1,0000	-
	10	0,5000	0,5000	1,0000	-
	15	0,5000	0,5000	1,0000	-
	20	0,5000	0,5000	1,0000	-
	25	0,5000	0,5000	1,0000	-
	30	0,5000	0,5000	1,0000	-
	35	0,5000	0,5000	1,0000	-
	40	0,5000	0,5000	1,0000	-
	45	0,5000	0,5000	1,0000	-
	50	0,5000	0,5000	1,0000	-

6.2.2 Distribuição Gama

Tabela 6.6: Estudo de simulação de classificação para intervalos de $C'_{pk}s$ para distribuição Gama.

Cenário	n	MVM				MV			
		AC	VPP	SB	EP	AC	VPP	SB	EP
Fronteira	5	0,5021	0,9565	0,0044	0,9956	0,5009	0,9474	0,0018	0,9982
	10	0,5166	0,8805	0,0383	0,9629	0,5134	0,8829	0,0031	0,9699
	15	0,5266	0,8323	0,0665	0,9369	0,5199	0,8287	0,0503	0,9516
	20	0,5400	0,8577	0,0958	0,9113	0,5282	0,8446	0,0690	0,9347
	25	0,5426	0,8556	0,1025	0,9055	0,5304	0,8462	0,0743	0,9299
	30	0,5455	0,8570	0,1091	0,8999	0,5306	0,8398	0,0755	0,9288
	35	0,5471	0,8589	0,1127	0,8970	0,5311	0,8410	0,0767	0,9278
	40	0,5477	0,8810	0,1103	0,8993	0,5312	0,8545	0,0752	0,9292
	45	0,5435	0,8559	0,1045	0,9038	0,5277	0,8411	0,0683	0,9353
	50	0,5444	0,8718	0,1040	0,9045	0,5280	0,8498	0,0679	0,9353
Ótimo	5	0,9995	1,0000	0,9991	0,5002	0,9987	1,0000	0,9974	0,5007
	10	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	15	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	20	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	25	0,9995	0,9998	1,0000	0,4999	0,9999	0,9980	1,0000	0,4999
	30	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	35	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	40	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	45	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	50	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
Descontrolado	5	0,4643	0,2580	0,0381	0,9589	0,4785	0,2655	0,0244	0,9745
	10	0,3235	0,3707	0,5063	0,2173	0,3231	0,3696	0,5014	0,2240
	15	0,4498	0,5000	1,0000	0,000	0,4520	0,4748	0,9033	0,0008
	20	0,4953	0,5000	1,0000	0,000	0,4956	0,4978	0,9912	0,000
	25	0,4998	0,5000	1,0000	0,000	0,4998	0,4999	0,9996	0,000
	30	0,5000	0,5000	0,9999	0,000	0,4998	0,4999	0,9999	0,000
	35	1,0000	0,5000	1,0000	0,5000	0,5000	0,5000	1,0000	0,000
	40	1,0000	0,5000	1,0000	0,5000	0,5000	0,5000	1,0000	0,000
	45	1,0000	0,5000	1,0000	0,5000	0,5000	0,5000	1,0000	0,000
	50	1,0000	0,5000	1,0000	0,5000	0,5000	0,5000	1,0000	0,000

Para o método de estimação MVM, é possível verificar que para o cenário fronteira os índices controlados foram classificados corretamente em torno de 90% das vezes, como pode-se ver pelo valor de especificidade, em contrapartida os itens descontrolados não. O que levou a uma acurácia em torno de 0,5. Para o cenário ótimo, o modelo foi capaz de classificar corretamente o descontrole, diferente da classificação pontual . Para o último cenário, o descontrolado, atingiu uma acurácia de 50%, e o modelo identificou verdadeiro negativos, índices controlados, uma vez que o processo não deveria tê-los, o mesmo ocorre para a estimação indireta, MV.

Ao comparar ambos vemos que, no geral, não há grandes perdas em utilizar a estimação direta, uma vez que o comportamento das métricas em ambos são parecidos. Observe o valor de sensibilidade, mesmo utilizando o método de estimação direta a maior parte dos itens descontrolados são identificados. Portanto, mesmo utilizando o método de estimação direta para a Gama, MVM, o índice identifica os itens descontrolados assim que ocorrem.

6.2.3 Distribuição Weibull

Tabela 6.7: Estudo de simulação de classificação para intervalos de $C'_{pk}s$ para distribuição Weibull.

Cenário	n	MLM				MV			
		AC	VPP	SB	EP	AC	VPP	SB	EP
Fronteira	5	0,4991	0,3208	0,0017	0,9983	0,5390	0,5657	0,3353	0,6889
	10	0,5061	0,5072	0,4294	0,5258	0,6596	0,9022	0,3580	0,7286
	15	0,4658	0,4627	0,4241	0,5447	0,8683	0,8982	0,8306	0,5217
	20	0,4352	0,4281	0,3858	0,5598	0,9151	0,8833	0,9565	0,4774
	25	0,4092	0,3990	0,3591	0,5612	0,9108	0,8580	0,9845	0,4595
	30	0,3831	0,3661	0,3197	0,5827	0,8993	0,8352	0,9949	0,4468
	35	0,3692	0,3458	0,2935	0,6025	0,8870	0,8170	0,9974	0,43 78
	40	0,3543	0,3224	0,2645	0,6267	0,8829	0,8111	0,9985	0,4346
	45	0,5435	0,8559	0,1045	0,9038	0,8761	0,8023	0,9980	0,4304
	50	0,3379	0,2915	0,2267	0,6645	0,8699	0,7936	0,9996	0,4254
Ótimo	5	0,9999	0,9999	1,0000	0,4999	1,0000	1,0000	1,0000	0,5000
	10	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	15	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	20	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	25	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	30	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	35	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	40	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	45	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
	50	1,0000	1,0000	1,0000	0,5000	1,0000	1,0000	1,0000	0,5000
Descontrolado	5	0,5000	0,5000	1,0000	-	0,4901	0,4950	0,9801	-
	10	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	15	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	20	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	25	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	30	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	35	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	40	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	45	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-
	50	0,5000	0,5000	1,0000	-	0,5000	0,5000	1,0000	-

Para o método de estimação MLM, um processo com distribuição Weibull e que tem $C_{pk} = 1.01$, o modelo performa melhor para amostras pequenas, de tamanho até 10. Porém, a sensibilidade, que representa a proporção de itens descontrolados classificados corretamente ainda é menor que 50%, por outro lado, os classificados como controlados corretamente são maiores, talvez se deve ao fato do índice estimado estar na fronteira entre o capaz e não capaz. Para o cenário ótimo, o descontrole causado foi identificado corretamente. Para o cenário descontrolado, ocorre o mesmo, porém nenhum índice controlado identificado, era o que se esperava,uma vez que todo o processo não é capaz.

Olhando exclusivamente para o método MV, ao aumentar o tamanho amostral a medida de acurácia também aumenta nos cenário fronteira e descontrolado.

Sendo assim, não há grandes perdas, quando estamos utilizando amostras pequenas, em utilizar a estimação direta quando o desejo maior é reduzir o tempo computacional.

6.2.4 Comparação entre as simulações pontuais e intervalares

Nesta seção, alguns pontos são levantados sobre a comparação entre a simulação intervalar e pontual dos índices de capacidade de processo C_{pk}.

6.2.4.1 Distribuição Normal

Ao comparar os resultados pontuais com os intervalares notamos que para o cenário ótimo e fronteira as métricas aumentaram ligeiramente. Por sua vez, em nenhum dos casos o modelo com cenário fronteira foi capaz de identificar itens controlados, uma explicação plausível é que o valor do índice está no limite do que é dito capaz e não capaz, ou seja, a estimação pode estar subestimada e resultar apenas em índices abaixo de um. Desta forma, fica o alerta para processos nos quais o índice de C_{pk} é próximo de 1,01, uma vez que o processo de estimação, nos casos testados, subestimou estes índices.

6.2.4.2 Distribuição Gama

Comparando os resultados intervalares e pontuais para a distribuição Gama, observamos que para os cenários fronteira e ótimo, a estimação para os intervalos resultou em melhores resultados na classificação dos itens descontrolados, apresentada pela medida VPP, e também para os controlados, principalmente no primeiro cenário citado, o que também elevou a medida de proporção de acertos. Em contrapartida, para os valores intervalares itens descontrolados foram considerados controlados, no cenário descontrole, o que era esperado, uma vez que o intervalo pode ser considerado capaz, mas a medida pontual não, e por isto a visualização gráfica se torna importante.

6.2.4.3 Distribuição Weibull

Analisando os resultados para as estimações pontuais e intervalares, é possível notar que para o cenário ótimo, independente do método de estimação todas medidas aumentaram. Para o cenário de descontrole, o comportamento é parecido, tendo um ganho, nas medidas de classificação, para o método de estimação de verossimilhança. Para o cenário fronteira, a especificidade para os intervalos foi maior quando comparado ao pontual, ou seja, foi possível de identificar os controlados corretamente, considerado um resultado melhor.

Capítulo

7

Aplicação

Neste capítulo, com o objetivo de ilustrar a ideia proposta, que é verificar de utilizar um método visual, rápido e prático, que discrimina amostras descontroladas em um processo em tempo real utilizando estimação direta usamos informações sobre um processo de ensacamento de achocolatado em pó.

Os dados foram obtidos em dois dias, coletados em cada dia por um operador mas seguindo sempre o mesmo procedimento de coleta, que consistiu em registrar os dados, calcular a média de cada amostra e realizar o gráfico para melhor visualização do comportamento no processo, além de registrar todo e qualquer tipo de ocorrência durante o processo.

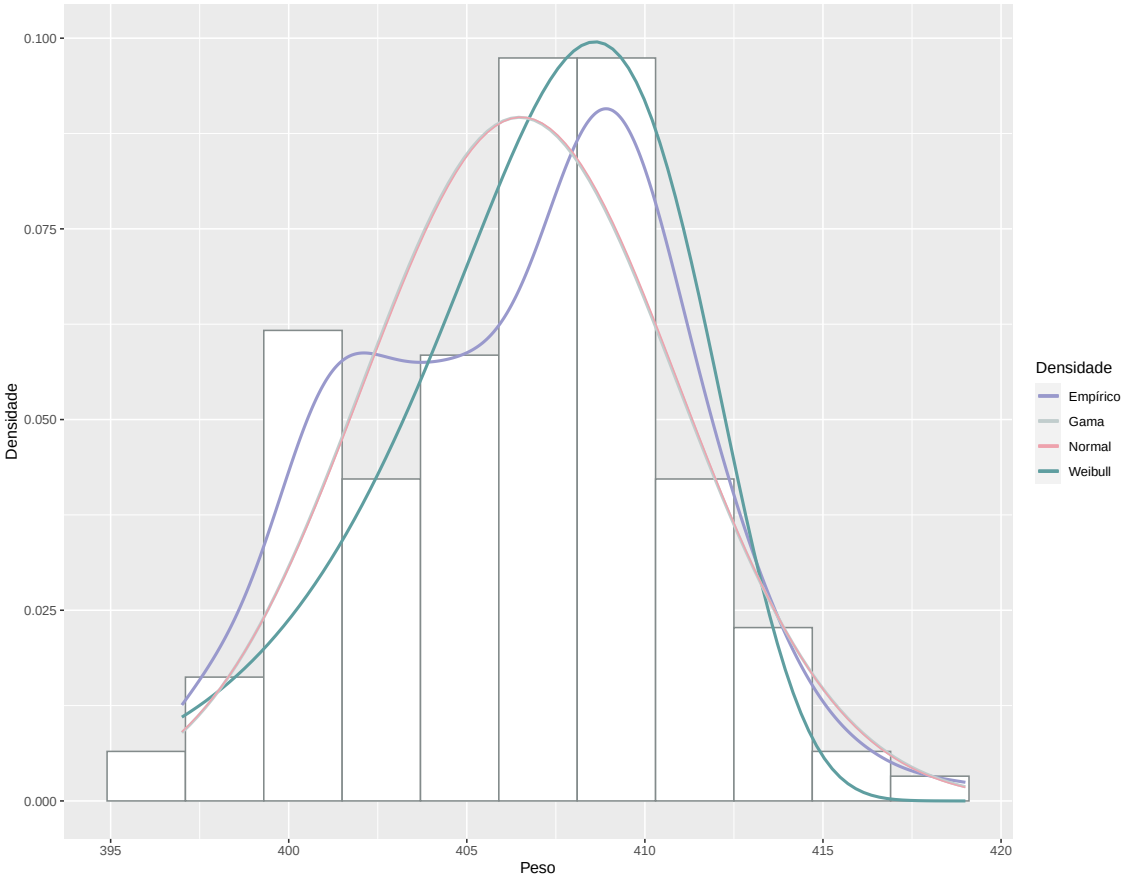
Cada embalagem deve ter 400 gramas, e ainda, segundo a portaria INMETRO nº 248 de 17/07/2008, a tolerância para produtos com conteúdo líquido pré-medidos de 400 gramas é de 3%, ou seja, os limites de especificação para o peso do achocolatado são de 388 e 412.

O conjunto de dados é constituído por 145 sacos de achocolatado coletados em 29 amostras de tamanho cinco, que são apresentados na Tabela 7.1 .

Tabela 7.1: Dados sobre o peso das embalagens de achocolatado.

	Subamostras									
	1	2	3	4	5	1	2	3	4	5
Amostras	403	414	412	408	416	410	405	408	413	408
	410	408	410	408	402	410	398	405	403	409
	408	412	405	414	410	412	409	413	405	408
	404	401	401	401	410	412	409	410	410	409
	401	405	403	408	405	408	411	403	397	404
	409	401	408	401	401	398	408	414	402	401
	402	404	410	404	400	401	412	409	416	409
	401	403	412	405	406	407	401	402	397	401
	402	408	410	407	404	411	406	413	408	401
	407	**409**	**410**	**397**	**406**	404	407	408	407	410
	405	409	406	410	419	409	410	406	408	410
	404	411	412	410	411	400	402	404	409	399
	407	409	411	404	410	398	403	398	407	400
	405	408	406	409	409	411	402	401	406	409
	408	401	406	400	413					

Figura 7.1: Histograma para as curvas de densidade calculadas segundo os parâmetros estimados do conjunto de dados centralizado na média.



Ao realizar as análises de controle para o processo identificamos que a amostra de número dez infringe a regra de Shewart, sendo assim, utilizamos os dados sem a amostra de número dez, em negrito, para encontramos qual distribuição, Normal, Gama ou Weibull, pode descrever estes dados utilizando as estimativas na forma direta apresentadas nos capítulos 3 e 4. Para isto vamos realizar um histograma com a densidade empírica e de cada distribuição para vermos o que ocorre.

Aparentemente, a distribuição Weibull se ajusta melhor à estes dados quando comparado com a distribuição Gama, para eliminarmos eventuais dúvidas realizaremos o teste de Kolmogorov-Smirnov (KS). Abaixo temos uma tabela com os parâmetros calculados pelos métodos diretos de cada distribuição e os p-valores resultantes do teste KS.

Tabela 7.2: Resultados do teste e parâmetros testados com base nos dados.

Distribuição Testada	Parâmetros Estimados		p-valor do Teste KS
Normal	406,5286	4,4504	0,0182
Gama	0,0487	8340,3530	0,0167
Weibull	408,6295	110,5432	0,1528

O teste numérico confirma as impressões apresentadas ao gráfico para o nível de confiança de 95%, ou seja, os dados podem ser representados pela distribuição Weibull com os parâmetros de escala e forma dados por 408,6295 e 110,5432, respectivamente, lembrando que estes foram calculados utilizando o método de estimação L-Momentos, que definimos como um método de estimação direta.

Definida a distribuição, vamos fazer a análise de controle de qualidade.
Toda a aplicação foi realizada no software livre R.

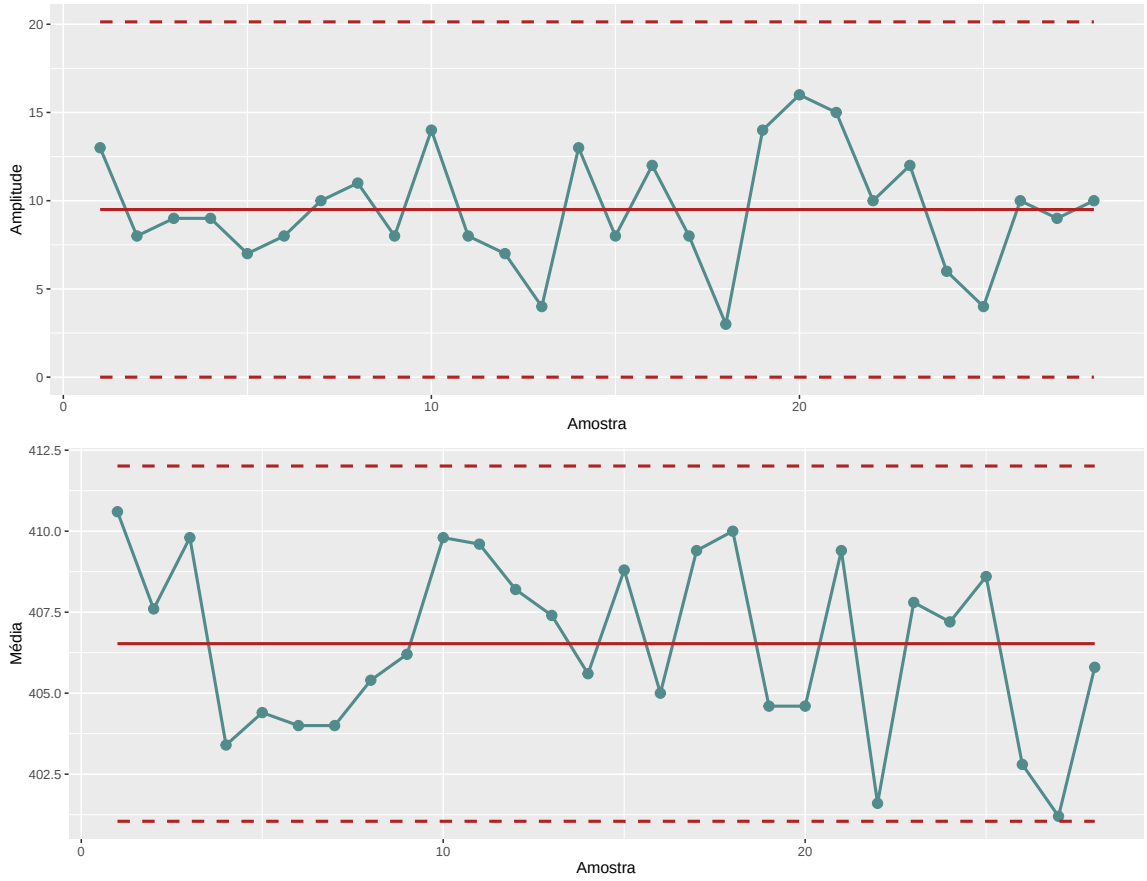
7.1 Análise de capacidade do processo.

Ao realizar as cartas de controle para os dados apresentados na Tabela 7.1, considerando que temos vinte e nove amostras de tamanho cinco, observamos que a amostra de número dez infringia uma das leis de Shewhart. Com isto, retiramos está amostra e revisemos os cálculos, e encontramos os novos limites superior e inferior para as cartas, como também a média das médias e amplitude média dada pela Tabela 7.3.

Tabela 7.3: Limites de Controle para as cartas de amplitude e média

Limites de Controle	Carta para amplitude	Carta para média
Inferior	0,000	401,049
Central	9,500	406,529
Superior	20,086	412,008

Figura 7.2: Cartas para amplitude e média dos pesos dos sacos de achocolatado.



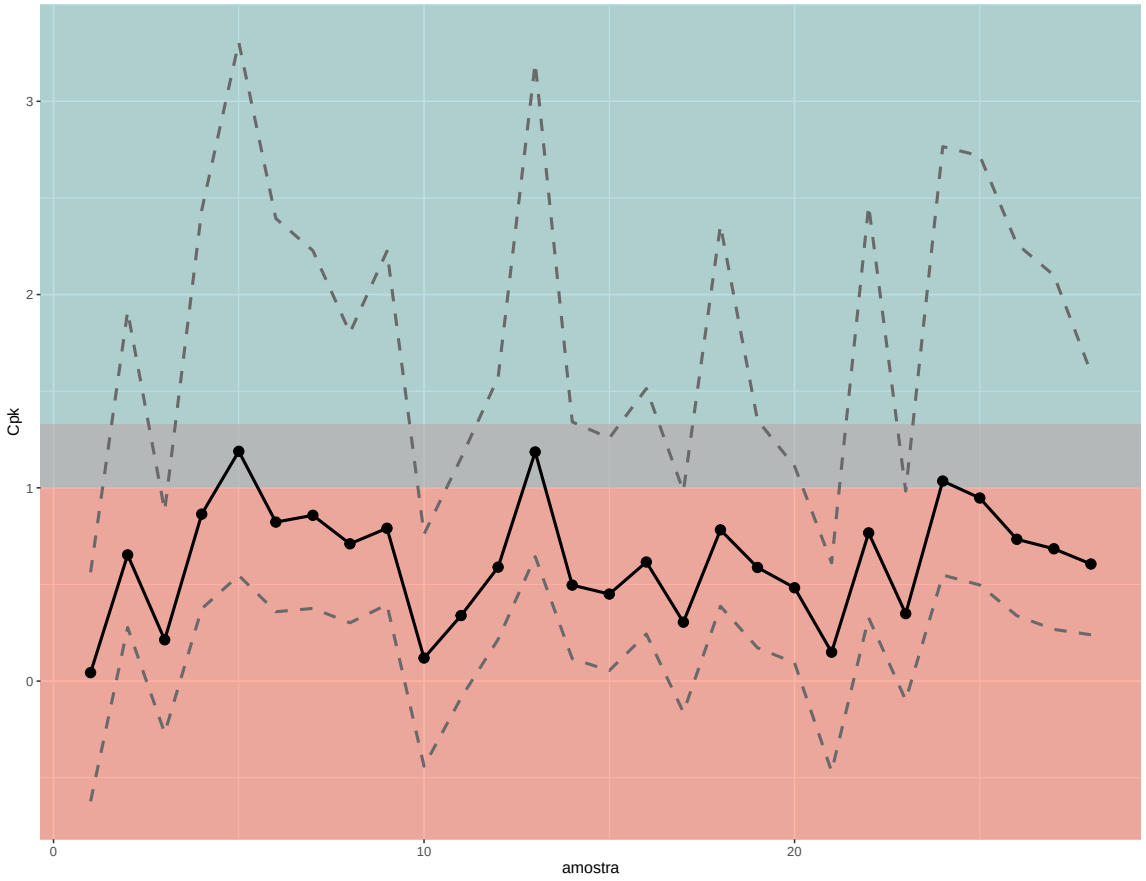
Ao observar as cartas observamos que o processo não está centrado na média nominal, no caso, quatrocentos gramas, e ainda, em média 6,5286 gramas a mais são colocadas em cada embalagem.

Apesar disto, não há pontos fora dos limites de controle, ou seja, aparentemente não há nada de errado durante o processo de ensacamento do achocolatado, após a retirada da amostra de número dez. Verificamos então se este processo respeita os limites de

especificação inferior dados pelo INMETRO, por sua vez ensaca 0,0110 gramas à mais que a legislação diz, ou seja, a empresa perde dinheiro ensacando uma quantidade maior, extrapolando até mesmo a recomendação do INMETRO.

Por outro, lado será que este processo é capaz? Para isto, vamos utilizar o gráfico criado pela fórmula *cpk.plot()*, descrita no Capítulo 5.

Figura 7.3: Gráfico C_{pk} para as amostras dos pesos dos achocolatados.



Observe que ao analisarmos os valores pontuais, apenas em três amostras o processo é considerado capaz , mas ao olhar para os intervalos de 95% de confiança vemos que apenas quatro amostras das 28 amostras são consideradas incapazes. Os valores dos índices calculados assim como seus intervalos são apresentados no Apêndice C, na Tabela C.1.

Futuramente, seria interessante estudar a interpretação intervalar para valores de intervalos menores, ou seja, com um nível de confiança maior que 95%.

7.2 Aplicação em tempo real.

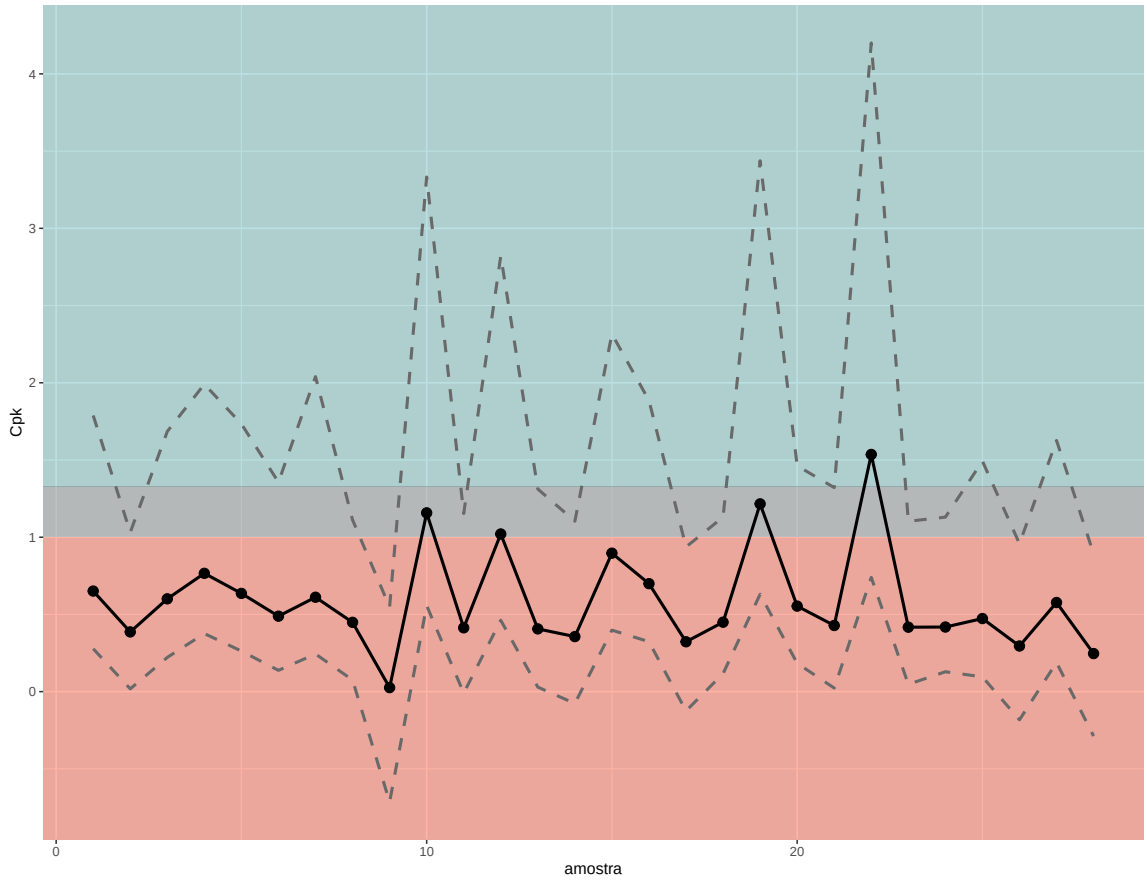
Nesta seção, usaremos os parâmetros encontrados pela estimação direta dos dados apresentados na Tabela 7.1 para simular um processo com tamanho de amostra igual ao anterior, e acrescentar uma pertubação em parte dos dados e verificar como nosso método identifica o descontrole causado, propositalmente, nas amostras.

Os dados gerados são apresentados em Apêndice C na Tabela C.4. O transtorno causada foi entre as observações 70 e 100, ou seja, as amostras 14 e 20 possuem acréscimo de cinco unidades.

Afim de identificar como o C_{pk}, calculado a partir de um método de estimação direta, capta a pertubação, recorrermos ao gráfico resultante da função *cpk.plot()*. Na Figura

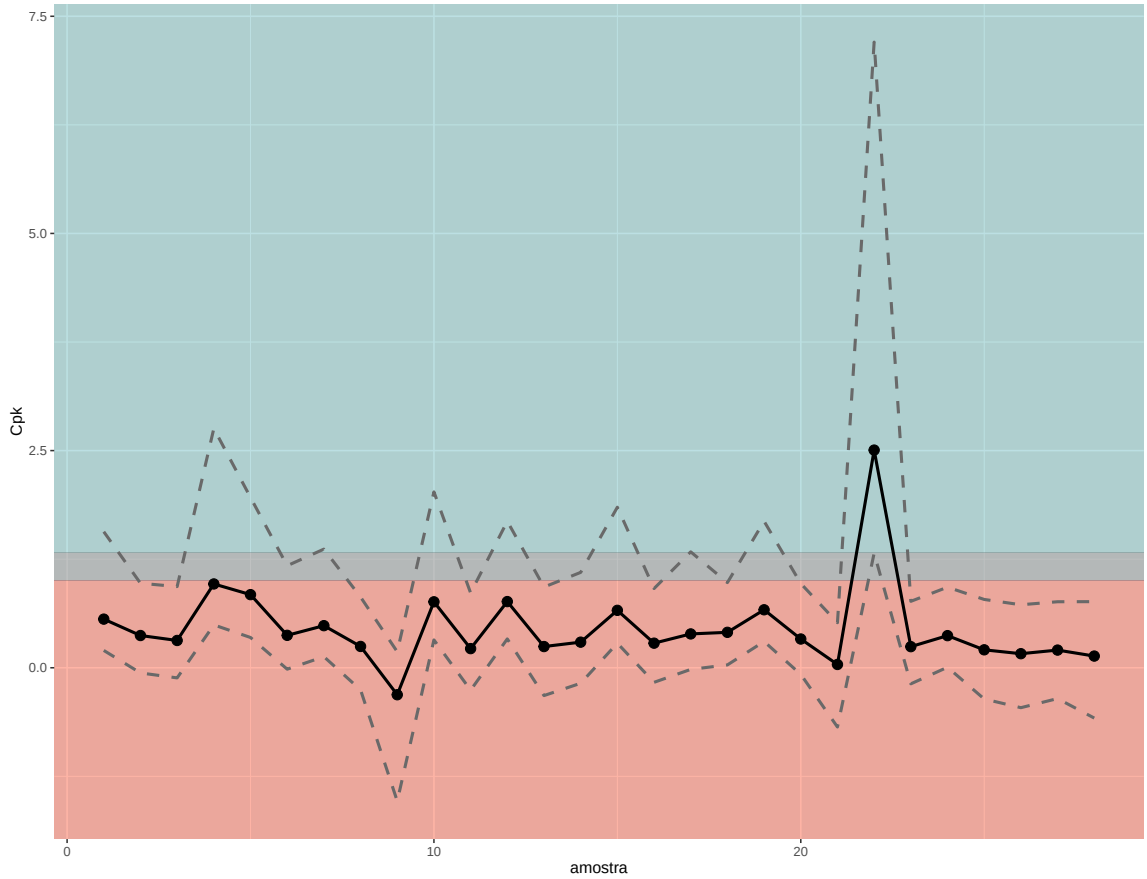
7.4 apresentamos o método para o processo integro, sem o acréscimo das unidades, já na Figura 7.5 o processo transtornado. As tabelas estão localizadas no Apêndice C.

Figura 7.4: Valores de C_{pk} para a amostra sem pertubação.



Observe que para a amostra perturbada, até interpretação intervalar indica que estas não são capazes de atentar às especificações, já para os dados que não tiveram o transtorno uma quantidade maior de amostras são consideradas capazes.

Figura 7.5: Valores de C_{pk} para a amostra com pertubação.



Portanto, o método exposto neste funciona, ou seja, o índice C_{pk} de Clemensts [2], calculado a partir de um método de estimação direta é capaz de identificar mudanças em tempo real no processo.

Capítulo

8

Considerações e Sugestões Futuras

Como visto, avaliar processos é de suma importância para gerar melhores resultados, diminuir custos e aumentar a competitividade. Desta forma, esta dissertação trás este olhar para a análise de desempenho de um processo de forma prática, rápida e em tempo real.

Sendo assim, utilizamos os métodos de estimações diretas para as distribuições Normal, Gama e Weibull, para otimizar o tempo operacional, sendo estes métodos, em média, vinte vezes mais rápidos quando comparado com o método de estimação indireta, discutida aqui. E como visto na simulação não há grandes diferenças para o valor de acurácia entre os métodos, de forma fechada e não linear.

Resumidamente, podemos dizer que para fins práticos visando a economia de tempo os métodos de estimação direta funcionam, cada um com sua particularidade, o método de estimação L-Momentos é robusto para amostras menores quando comparado ao método de Verossimilhança, e o método de verossimilhança modificado se assemelha ao desempenho do Método de verossimilhança, ou seja, quanto maior o tamanho amostral melhor os resultados obtidos.

Durante a simulação e aplicação podemos concluir que para valores intervalares de C_{pk} o método proposto funciona, ou seja, calcular o intervalo para o índice utilizando um método de estimação direta em um processo, em tempo real, quando existe uma pertubação o índice irá detectar, sendo ele utilizando a estimação direta ou não.

Uma simples sugestão para estudos futuros seria acrescentar novas distribuições a função. Deixa-lá mais robusta, incluindo testes para identificação da distribuição. Também estender a análise para outros índices de capacidade, como por exemplo o C_{pm}, que considera tanto a média e a variabilidade do processo como também a média da especificação nominal. Tornando a ferramenta gráfica mais robusta, e também como uma ótima alternativa para o auxílio na tomada de decisões.

Uma possível alternativa, seria automatizar todo o processo, deste a coleta dos dados, passando pela estabilização do processo e chegando até a análise gráfica dos índices, incorporar a ideia desta dissertação aos processos automatizados, facilitando a tomada de decisão no tempo real.

Referências

[1] G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[2] J A Clements. Process capability calculations, for non-normal distributions. *Quality progress*, 22(9):95 – 100, 1989.

[3] Heleno Bolfarine e Mônica Carneiro Sandoval. *Introdução à Inferência Estatística*. SBM, 2010.

[4] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

[5] Elsayed A.H. Elamir and Allan H. Seheult. Trimmed l-moments. *Computational Statistics & Data Analysis*, 43(3):299–314, 2003.

[6] J. Arthur Greenwood, J. Maciunas Landwehr, N. C. Matalas, and J. R. Wallis. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 15(5):1049–1054, 1979.

[7] J. R. M. Hosking. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1):105–124, 1990.

[8] Z. Hosseinifard, B. Abbasi, and S.T.A. Niaki. Process capability estimation for leukocyte filtering process in blood service: A comparison study. *IIE Transactions on Healthcare Systems Engineering*, 4(4):167–177, 2014.

[9] N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2):149–176, 1949.

[10] S Kotz and R. Lovelace, C. *Process Capability Indices in Theory and Practice*. Arnold Publishers Oxford University Press, 1998.

[11] Francisco Louzada, Carlos Diniz, Paulo Ferreira, and Edil Ferreira. *Controle estatístico de processos: uma abordagem prática para cursos de engenharia e administração*. Grupo Gen-LTC, 2013.

[12] Francisco Louzada, Pedro Ramos, and Eduardo Ramos. A note on bias of closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician*, 73(2):195 – 199, 2019.

[13] Francisco Louzada, Pedro L Ramos, and Gleici SC Perdoná. Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data. *Computational and Mathematical Methods in Medicine*, 2016.

[14] P.L. Meyer. *PROBABILIDADE Aplicações à Estatística.* LTC - Livros Técnicos e Científicos Editora, 1983.

[15] DOUGLAS C MONTGOMERY. Introdução ao controle estatístico da qualidade. 4ª edição. *Rio de Janeiro: Editora LTC*, pages 95–108, 2004.

[16] Manuel Pina-Monarrez, Manuel Rodríguez-Borbón, and Jesus Ortiz-Yañez. Non-normal capability indices for the weibull and lognormal distributions. *Quality and Reliability Engineering*, 2016.

[17] Pedro L. Ramos, Eduardo Ramos, Francisco A. Rodrigues, and Francisco Louzada. A modified closed-form maximum likelihood estimator, 2021.

[18] LAR. Rivera, N. F Hubele, and F.P Lawrence. Cpk index estimation using data transformation. *Comput Ind Eng*, 29:55–58, 1995.

[19] Bahar Sennaroglu and Ozlem Senvar. Performance comparison of box-cox transformation and weighted variance methods with weibull distribution. *Journal of Aeronautics and Space Technologies (Havacilik ve Uzay Teknolojileri Dergisi)*, 8, 2015.

[20] Ozlem Senvar and Cengiz Kahraman. Type-2 fuzzy process capability indices for non-normal processes. *Journal of Intelligent & Fuzzy*, 01 2014.

[21] S.E. Somerville and D.C. Montgomery. Process capability indices and non-normal distributions. *Quality Engineering*, 19:305–316, 1996-1997.

[22] Fred Spiring, Bartholomew Leung, Smiley Cheng, and Anthony Yeung. A bibliography of process capability papers. *Quality and Reliability Engineering International*, 19(5):445–460, 2003.

[23] Mahdi Teimouri, Seyed M. Hoseini, and Saralees Nadarajah. Comparison of estimation methods for the weibull distribution. *Statistics*, 47(1):93–109, 2013.

Apêndice

A

Estudo de simulação para os índices C_{pk}

A seguir, apresenta-se resultados do estudo de simulação para os valores do índice C_{pk} para distribuição e cenário descritos no capítulo 6.

As medidas apresentadas foram:

$$\begin{aligned} EQM_{C_{pk}} &= \sum_{i=1}^{n} \frac{(\hat{C_{pk}}_i - C_{pk})^2}{n} \\ EMR_{C_{pk}} &= \sum_{i=1}^{n} \frac{\hat{C_{pk}}_i}{nC_{pk}}, \end{aligned} \tag{A.1}$$

sendo n o tamanho amostral, $\hat{C_{pk}}$ o índice estimado baseado nas estimativas dos parâmetros, e por fim C_{pk} o valor índice real, calculado com os parâmetros reais, ou seja, aqueles que serão estimados.

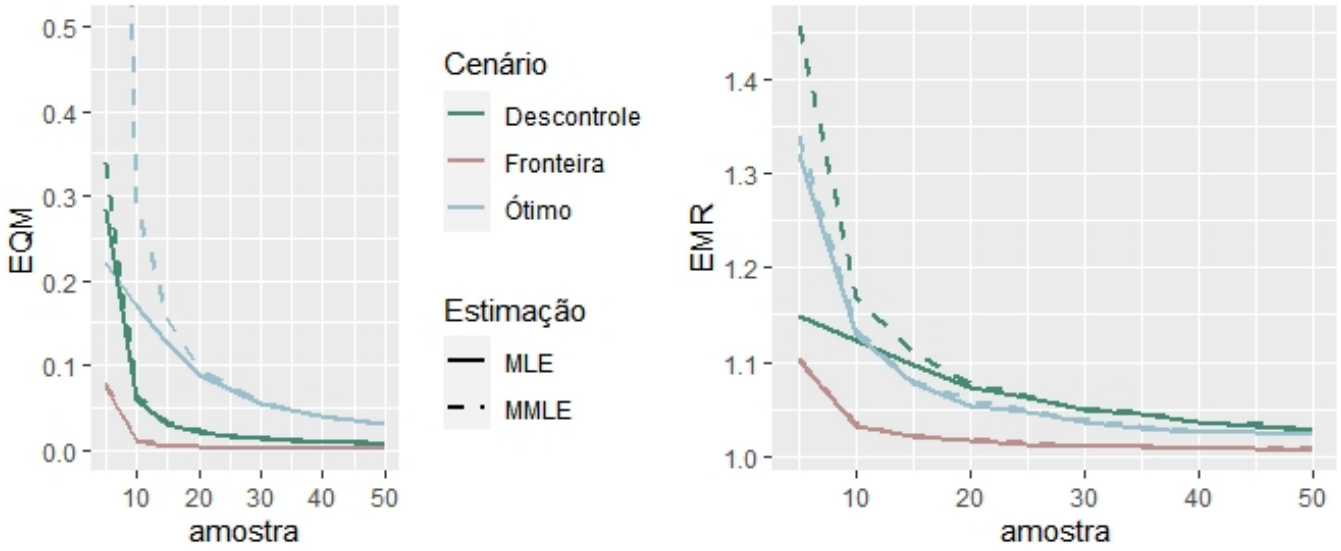
Aqui, foram geradas 10000 simulações, nas quais usou-se $n = 5, 10, 15, 20, 25, 30, 35, 40, 45$ e 50. Nas próximas seções são apresentados breves resultados para as distribuições trabalhadas neste.

A.1 Distribuição Gama

Retomando os cenários para distribuição Gama temos:

- Cenário ótimo: com $C_{pk} = 1,33$, com $\lambda = 50$, $\phi = 5$ e limites de especificação inferior e superior iguais a 0 e 16,37847;
- Cenário fronteira: sendo $\lambda = 2$, $\phi = 1$, $LIE = 0,03$ e $LSE = 12,5$, com $C_{pk} = 1,014078$;
- Cenário descontrole: $C_{pk} = 0,760837$, gerado pelos parâmetros $\lambda = 55, \phi = 10$ e limites de especificação $LIE = 4$ e $LSE = 16,37847$.

Figura A.1: EQM e EMR para os índices C_{pk} calculados segundo distribuição Gama.



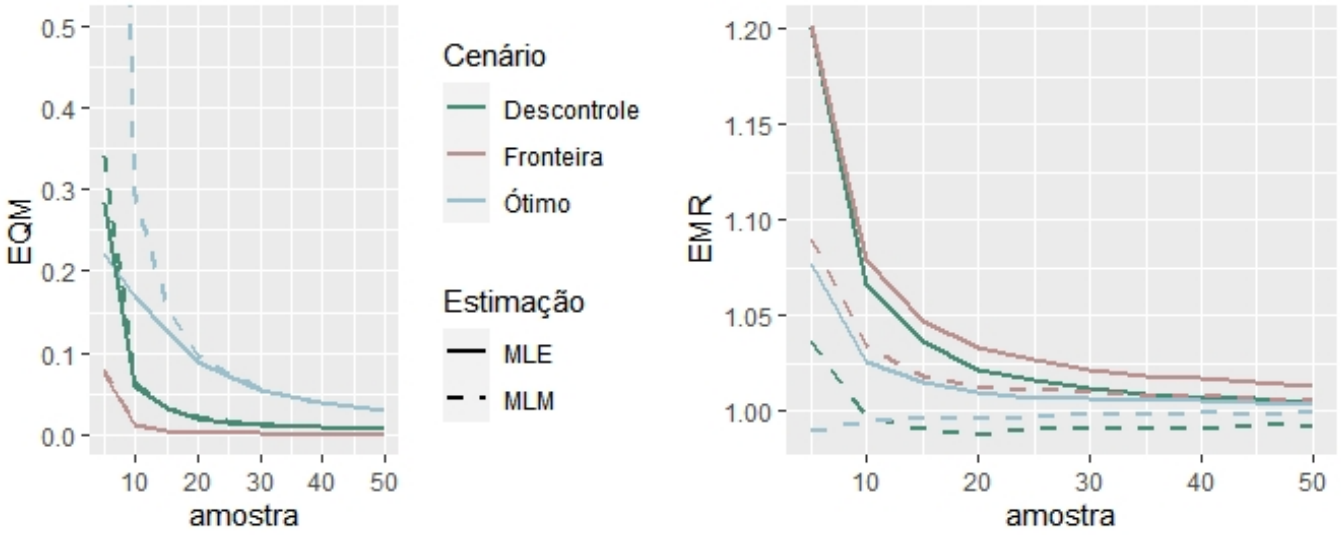
Observe que independente do cenário: ao aumentar o tamanho amostral o comportamento para os dois métodos de estimações é bem parecido, ou seja, não há grandes percas utilizando o método de verossimilhança modificado, MMLE, pelo contrário existe o ganho do tempo, uma vez que este processo consome menos recursos computacionais.

A.2 Distribuição Weibull

Antes de apresentar o breve resultado para cada cenário, relembramos :

- Cenário ótimo: $\alpha = 5, \beta = 3, LIE = 0.1443575$ e $LSE = 5$, o que dá $C_{pk} = 1.33$
- Cenário fronteira: $C_{pk} = 1,01$, gerado por $\alpha = 4, \beta = 2, LIE = 0.369016$ e $LSE = 10$;
- Cenário descontrole: $\alpha = 2, \beta = 0,5, LIE = 0,14$ e $LSE = 2$, resultando $C_{pk} = 0,6943387$.

Figura A.2: EQM e EMR para os índices C_{pk} calculados segundo distribuição Weibull.



Observe que para todos os cenários, o método MLM produz valores de Erro Médio Relativo (EMR) melhores quando comparado com o método de verossimilhança, ou seja, são mais próximos de um. E os valores de erro quadrático médio para os cenários descontrole e fronteira são bem próximos, e para amostras maiores de 20 são quase iguais, independentemente do cenário.

Apêndice

B

Código para a função *cpk.plot*

```
cpk.plot = function(data,n,m,LSE,LIE,dist){
library("ggplot2");library("gridExtra");library("qcc")
Cpk<-c(); CpkSup<-c();CpkInf<-c(); CpkB<-c()
amostra= 1:m;B =1000
data <- qcc.groups(data,rep(amostra,n))
if(dist == "Normal"){
for(i in 1:m){ w<-c() for(j in 1:n){w[j]<-data[i,j] }
nfit = fitdistr(w,"normal");media = nfit$estimate[1]
dp =nfit$estimate[2];
cpknorm = min((LSE - media)/(3*dp),(LIE - media)/(3*dp))
Cpk[i]<-cpknorm
for(k in 1:B){l = rnorm(n,media ,dp);nfitB = fitdistr(l,"normal")
mediaB = nfitB$estimate[1]; dpB = nfitB$estimate[2]
CpkB[k]= min((LSE - mediaB)/(3*dpB),(LIE - mediaB)/(3*dpB))}
CpkSup[i]<-quantile(CpkB, probs = 0.975, na.rm = FALSE,names = FALSE,type = 7)
CpkInf[i]<- quantile(CpkB, probs = 0.025, na.rm = FALSE,names = FALSE,type = 7)}
Cpk = as.data.frame(cbind(amostra,CpkInf,Cpk,CpkSup));inferior<-min(CpkInf);
superior<-max(CpkSup);gr = ggplot(Cpk, aes(x=amostra,y=Cpk))+
scale_y_continuous(limits = c(min(Cpk,inferior),max(1.6,superior))) +
geom_rect(xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = 1,fill = "tomato",
alpha = 0.02)+geom_rect(xmin = -Inf, xmax = Inf, ymin = 1.33, ymax = Inf,
fill = "cadetblue3", alpha = 0.02)+geom_rect(xmin = -Inf, xmax = Inf,
ymin = 1, ymax = 1.33, fill = "azure4", alpha = 0.02)+geom_line(size=1)
+geom_point(size=3)+geom_line(aes(y = CpkInf), linetype=2,size=1,
col="dimgray")+geom_line(aes(y = CpkSup), linetype=2,size=1,col="dimgray")
op = print(gr);tab = Cpk[,-1];tabela=cbind(tab);colnames(tabela)=
c("I.C Inferior","Cpk - MV","I.C Superior")
return(tabela);par(op) }
if (dist=="Gamma"){for(i in 1:m){x<-c();for(j in 1:n){x[j]<-data[i,j]}
if (dist=="Gamma"){for(i in 1:m){x<-c();for(j in 1:n){x[j]<-data[i,j]}
```

```
hbet<-(1/(n*(n-1)))*(n*sum((x)*log(x))-s2*sum(log(x)))
Upv <- qgamma(0.99865,shape=halpha,scale=hbet);Lpv <- qgamma(0.00135,
shape=halpha,scale=hbet);Mv <- qgamma(0.5,shape=halpha,scale=hbet)
cpkgama <- min((LSE - Mv)/(Upv-Mv), (Mv - LIE)/(Mv-Lpv));Cpk[i]<-cpkgama
for(k in 1:B){z = rgamma(n,shape=halpha ,scale=hbet)
balphaM = (n*sum(z)) / (n*sum(z*log(z)) - (sum(z)*sum(log(z))))
bbetM=(1/(n^2))*(n*sum((z)*log(z))- (sum(z)*sum(log(z))))
UM=qgamma(0.99865,balphaM,scale=bbetM);LM <- qgamma(0.00135,
balphaM,scale=bbetM);MM=qgamma(0.5,balphaM,scale=bbetM)
cpkbootM<- min((LSE - MM)/(UM-MM), (MM - LIE)/(MM-LM));CpkB[k] <- cpkbootM}
CpkSup[i]<-quantile(CpkB, probs = 0.975, na.rm = FALSE,names = FALSE,type = 7)
CpkInf[i]<- quantile(CpkB, probs = 0.025, na.rm = FALSE,names = FALSE,type = 7)}
CpkD = as.data.frame(cbind(amostra,CpkInf,Cpk,CpkSup))
inferior<-min(CpkInf) ; superior<-max(CpkSup)
gr = ggplot(CpkD, aes(x=amostra,y=Cpk))+
scale_y_continuous(limits = c(min(Cpk,inferior),max(1.6,superior))) +
geom_rect(xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = 1,fill = "tomato",
alpha = 0.02)+geom_rect(xmin = -Inf, xmax = Inf, ymin = 1.33,
ymax = Inf, fill = "cadetblue3", alpha = 0.02) +geom_rect(xmin = -Inf,
xmax = Inf, ymin = 1, ymax = 1.33, fill = "azure4", alpha = 0.02) +
geom_line(size=1) + geom_point(size=3)+geom_line(aes(y = CpkInf),
linetype=2,size=1,col="dimgray")+geom_line(aes(y = CpkSup),
linetype=2,size=1,col="dimgray");op = print(gr);tab = CpkD[,-1]
tabela=cbind(tab);colnames(tabela)=c("Estimação Verossimilhança
Modificada-Cpk","Intervalo Superior","Intervalo Inferior")
return(tabela);par(op) }
if (dist=="Weibull"){for(i in 1:m){y<-c()
for(j in 1:n){y[j]<-data[i,j]};y=sort(y);pos=(1:1:length(y))
m1=mean(y);m2=try(((2/(n^2 - n))*(sum((pos - 1)*y)))-mean(y))
formW=try((- log(2)/( log( 1 -( m2 / m1)))))
escW = try((m1/gamma( 1 / formW+1)))
UM=qweibull(0.99865,formW,scale=escW);LM=qweibull(0.00135,formW,
scale=escW);MM=qweibull(0.5,formW,scale=escW)
cpkweibull=min((LSE - MM)/(UM-MM),(MM - LIE)/(MM-LM))
Cpk[i]=cpkweibull;for(k in 1:B){w=sort(rweibull(n,shape=formW,
scale=escW));m1y=mean(w);posy=(1:1:length(w))
m2y=try(((2/(n^2 - n))*(sum((pos - 1)*w ))) - mean(w))
balpha=try(( - log(2)/( log( 1 -( m2y / m1y) ) )))
bbeta=try((m1y/gamma(1 / balpha+1)));UM=qweibull(0.99865,balpha,
scale=bbeta);LM=qweibull(0.00135,balpha,scale=bbeta);MM=qweibull
(0.5,balpha,scale=bbeta);cpkbootM=min((LSE - MM)/(UM-MM),
(MM - LIE)/(MM-LM));CpkB[k]=cpkbootM}
CpkSup[i]<-quantile(CpkB, probs = 0.975, na.rm = FALSE,names = FALSE,type = 7)
CpkInf[i]<- quantile(CpkB, probs = 0.025, na.rm = FALSE,names = FALSE,type = 7)}
```

```
75 CpkD=as.data.frame(cbind(amostra,Cpk,CpkSup,CpkInf))
76 inferior=min(CpkInf);superior<-max(CpkSup)
77 gr = ggplot(CpkD,aes(x=amostra,y=Cpk))+scale_y_continuous
78 (limits = c(min(Cpk,inferior),max(1.6,superior)))+geom_rect(xmin=-Inf,
79 xmax = Inf, ymin = -Inf, ymax = 1,fill = "tomato", alpha = 0.02)+
80 geom_rect(xmin = -Inf,xmax=Inf,ymin = 1.33, ymax = Inf,fill="cadetblue3",
81 alpha = 0.02)+geom_rect(xmin = -Inf, xmax = Inf, ymin = 1, ymax = 1.33,
82 fill = "azure4", alpha = 0.02)+geom_line(size=1) +geom_point(size=3)+
83 geom_line(aes(y = CpkInf), linetype=2,size=1,col="dimgray")+geom_line
84 (aes(y = CpkSup), linetype=2,size=1,col="dimgray");op = print(gr)
85 tab = CpkD[,-1];tabela=cbind(tab);colnames(tabela)=c("I.C Inferior",
86 "Cpk - MVM","I.C Superior",Inferior");return(tabela);par(op)}
87 else print("Distribuição não Encontrada")
```

Apêndice

C

Tabelas extras

.

No apêndice C, tabelas como *output* da função gráfica, os dados gerados para aplicação e os valores de correção das cartas de controle são apresentados.

Tabela C.1: Output da função *cpk.plot* para os dados de ensacamento de achocolatado.

	I.C Inferior	Cpk - MLM	I.C Superior
1	-0.62	0.04	0.55
2	0.27	0.65	1.86
3	-0.39	0.21	0.88
4	0.39	0.86	2.48
5	0.54	1.19	3.34
6	0.38	0.82	2.19
7	0.41	0.86	2.24
8	0.30	0.71	1.82
9	0.38	0.79	2.39
10	-0.54	0.12	0.62
11	-0.12	0.34	1.17
12	0.23	0.59	1.48
13	0.64	1.19	3.41
14	0.13	0.50	1.26
15	0.05	0.45	1.43
16	0.23	0.62	1.44
17	-0.11	0.30	0.96
18	0.38	0.78	2.18
19	0.20	0.59	1.44
20	0.11	0.48	1.02
21	-0.40	0.15	0.79
22	0.34	0.77	2.31
23	-0.07	0.35	1.10
24	0.55	1.04	3.23
25	0.51	0.95	2.42
26	0.32	0.73	2.42
27	0.29	0.68	2.22
28	0.23	0.61	1.45

Tabela C.2: Valores de C_{pk} para a amostra sem pertubação - saída numérica.

	I.C Inferior	Cpk - MLM	I.C Superior
1	0.28	0.65	1.79
2	0.02	0.39	1.03
3	0.22	0.60	1.68
4	0.38	0.77	1.99
5	0.26	0.64	1.73
6	0.14	0.49	1.36
7	0.24	0.61	2.04
8	0.08	0.45	1.11
9	-0.71	0.03	0.55
10	0.56	1.16	3.33
11	-0.01	0.41	1.15
12	0.46	1.02	2.82
13	0.03	0.41	1.31
14	-0.08	0.36	1.10
15	0.40	0.90	2.32
16	0.32	0.70	1.89
17	-0.12	0.32	0.94
18	0.12	0.45	1.13
19	0.63	1.22	3.44
20	0.18	0.55	1.46
21	0.02	0.43	1.32
22	0.74	1.54	4.20
23	0.05	0.42	1.10
24	0.13	0.42	1.13
25	0.10	0.47	1.49
26	-0.18	0.30	0.96
27	0.19	0.58	1.63
28	-0.29	0.25	0.89

Tabela C.3: Valores de C_{pk} para a amostra com pertubação - saída numérica.

	I.C Inferior	Cpk - MLM	I.C Superior
1	0.20	0.56	1.57
2	-0.06	0.37	0.97
3	-0.12	0.31	0.93
4	0.49	0.96	2.75
5	0.35	0.84	1.96
6	-0.02	0.37	1.17
7	0.13	0.48	1.37
8	-0.25	0.25	0.81
9	-1.54	-0.31	0.18
10	0.32	0.76	2.02
11	-0.26	0.22	0.86
12	0.33	0.76	1.69
13	-0.32	0.24	0.93
14	-0.18	0.29	1.10
15	0.28	0.66	1.85
16	-0.17	0.28	0.91
17	-0.02	0.39	1.33
18	0.03	0.41	0.98
19	0.30	0.67	1.69
20	-0.08	0.33	0.97
21	-0.68	0.04	0.52
22	1.32	2.51	7.20
23	-0.19	0.24	0.76
24	0.01	0.37	0.93
25	-0.36	0.21	0.79
26	-0.46	0.16	0.73
27	-0.35	0.20	0.76
28	-0.58	0.13	0.76

Tabela C.4: Dados simulamos conforme a distribuição do peso das embalagens de achocolatado.

	Subamostras				
	1	2	3	4	5
Amostras	403,2183	401,2487	411,7286	410,7373	405,0455
	407,4237	396,2416	410,1652	377,6745	412,9132
	396,2574	406,3142	394,1112	400,5325	403,5006
	399,7352	413,4845	404,0348	406,8480	408,1685
	409,6500	411,3514	404,4922	405,4430	405,5651
	402,4258	408,7103	407,1968	407,3970	410,7837
	402,7004	406,2659	412,1243	403,2345	407,7810
	409,9663	412,1009	407,3628	408,9495	403,4481
	391,8646	402,7373	403,0763	407,5510	401,9185
	405,0142	400,7829	410,0159	409,3602	401,9606
	404,3324	404,0350	403,7098	411,8428	401,5331
	401,1015	405,5233	406,2297	409,7247	391,1730
	398,6092	412,6146	403,3689	415,7960	409,5018
	402,2238	413,7698	412,1928	408,6459	403,5709
	405,5614	405,0536	414,3762	411,0840	410,7458
	413,1493	405,1164	405,6437	392,4128	409,4036
	403,7705	401,3222	410,5490	411,0465	414,9205
	403,3001	412,9317	406,2832	407,4555	400,9592
	413,3271	406,1929	399,6559	410,1316	408,9106
	408,5744	409,4466	407,5402	403,3817	407,2244
	408,8832	408,5712	397,7135	410,9171	403,2498
	411,2235	405,5508	407,2608	411,2311	403,9297
	386,0162	408,0465	400,1080	405,8176	416,6346
	408,4878	401,6198	394,1488	407,8173	405,3437
	410,6582	402,2749	405,9335	400,8292	404,2697
	399,2715	408,9255	407,3534	406,6056	406,8231
	409,2199	413,2681	404,6291	410,5232	398,9666
	408,8048	413,2792	407,2150	414,4188	407,4357

Tabela C.5: Fatores de correção para construção dos gráficos $\overline{X}$ e R.

n	d_2	d_3	n	d_2	d_3
2	1,1296	0,8541	7	2,7074	0,8326
3	1,6918	0,8909	8	2,8501	0,8209
4	2,0535	0,8800	9	2,9677	0,8102
5	2,3248	0,8674	10	3,0737	0,7890
6	2,5404	0,8508	11	3,1696	0,7890