

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM
QUANTITATIVA**

**ANÁLISE DE DADOS DE VIGILÂNCIA
EPIDEMIOLÓGICA COM AUXÍLIO DE GRÁFICOS DE
CONTROLE ESTATÍSTICO**

MONOGRAFIA DE ESPECIALIZAÇÃO

Caroline Pafiadache da Silva

Santa Maria, RS, Brasil

2013

ANÁLISE DE DADOS DE VIGILÂNCIA EPIDEMIOLÓGICA COM AUXÍLIO DE GRÁFICOS DE CONTROLE ESTATÍSTICO

Caroline Pafiadache da Silva

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, do Centro de Ciências Naturais e Exatas da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

Orientadora: Profa. Dra. Roselaine Ruviaro Zanini

Santa Maria, RS, Brasil

2013

**Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Curso de Especialização em Estatística e Modelagem Quantitativa**

A Comissão Examinadora, abaixo assinada,
aprova a Monografia de Especialização

**ANÁLISE DE DADOS DE VIGILÂNCIA EPIDEMIOLÓGICA COM
AUXÍLIO DE GRÁFICOS DE CONTROLE ESTATÍSTICO**

elaborada por
Caroline Pafiadache da Silva

como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa.

COMISSÃO EXAMINADORA:

Roselaine Ruviano Zanini, Dra.
(Presidente/Orientadora)

Adriano Mendonça Souza, Dr. (UFSM)

Luciane Flores Jacobi, Dra. (UFSM)

Santa Maria, 30 de Janeiro de 2013.

Aos meus pais,
irmã
e ao meu noivo.

AGRADECIMENTOS

À Professora Dra. Roselaine Ruviano Zanini pela orientação, paciência, dedicação, respeito, apoio e ensinamentos que permitiram concluir este trabalho.

Aos professores do Curso de Especialização em Estatística e Modelagem Quantitativa: Dr. Adriano Mendonça Souza, Dra. Anaelena Bragança de Moraes, Dra. Janete Pereira Amador, Dra. Lorena Vicini e Dra. Luciane Flores Jacobi pelo comprometimento da partilha do conhecimento.

Aos professores da Banca Examinadora pela colaboração e sugestões.

À minha família pelo apoio e incentivo proporcionando, como sempre, um ambiente adequado aos estudos.

Ao Gabriel Piekala pela postura incondicional.

À Universidade Federal de Santa Maria pela oportunidade.

A todos os que emanam energia positiva e que de alguma forma contribuem com o meu crescimento intelectual.

RESUMO

Monografia de Especialização
Curso de Especialização em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria

ANÁLISE DE DADOS DE VIGILÂNCIA EPIDEMIOLÓGICA COM AUXÍLIO DE GRÁFICOS DE CONTROLE ESTATÍSTICO

Autora: Caroline Pafiadache da Silva
Orientadora: Roselaine Ruviaro Zanini
Data e Local da Defesa: Santa Maria, 30 de janeiro de 2013.

Os gráficos de controle estatístico desempenham um papel importante em processos de vigilância em saúde. No entanto, a implementação desses gráficos ainda não se tornou uma rotina em meio aos profissionais dessa área. O objetivo deste estudo é apresentar uma revisão do referencial teórico a cerca de alguns gráficos de controle usados na área da saúde, identificar os principais gráficos de controle estatísticos propostos na literatura e apresentar uma aplicação prática de três métodos estatísticos para detecção precoce de epidemias de doenças de notificação compulsória: os Gráficos \bar{x} e R (para medidas individuais e amplitude móvel), gráfico da Soma Acumulada (CUSUM tabular bilateral) e gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA). Para isso, foi realizado um estudo transversal com dados secundários provenientes do SINAN (Sistema de Informação de Agravos de Notificação) constituído pelos casos de Hepatite C, registrados no Rio Grande do Sul, no período de janeiro de 2007 a dezembro de 2011. O desempenho dos gráficos foi analisado e discutido comparativamente. Além disso, utilizou-se a metodologia de Box e Jenkins para obter um modelo que captasse o comportamento da série em estudo e que atendesse as presunções do uso de gráficos de controle. O modelo encontrado foi um SARIMA $(0,1,1) \times (1,0,0)_{12}$. Pode-se concluir que o gráfico EWMA é mais sensível que os demais na detecção de perturbações no processo podendo, portanto, complementar de forma eficaz investigações epidemiológicas.

Palavras-chave: Vigilância epidemiológica, doenças endêmicas, gráficos de controle estatísticos, modelo SARIMA.

ABSTRACT

Specialization Monograph
Course Specialization in Statistics and Quantitative Modeling
Federal University of Santa Maria

DATA ANALYSIS OF EPIDEMIOLOGICAL SURVEILLANCE WITH AID CONTROL OF STATISTICAL GRAPHS

Author: Caroline Pafiache da Silva
Advisor: Roselaine Ruviaro Zanini
Date and Place of Defense: Santa Maria, January 30th 2013.

The statistical control charts play an important role in processes of health surveillance. However, implementation of these charts has not yet become a routine amid the healthcare practitioners. This aim proposes the statistical control charts play an important role in processes of health surveillance. However, implementation of these charts has not yet become a routine amid the healthcare practitioners. This aim proposes two statistical methods for early detection of outbreaks of notifiable diseases: Graphs \bar{x} and R (for individual measurements and moving range) and exponentially weighted moving average (EWMA). For this, we performed a cross-sectional study using secondary data from the SINAN (Information System for Notifiable Diseases) consisting of cases of hepatitis C, reported in Rio Grande do Sul, in the period January 2007 to December 2011. The graphics performance were analyzed and discussed comparatively. For this purpose, we used the methodology of Box and Jenkins to obtain a model to captures the behavior of the series and to meet the presuppositions of the use of control charts. The model found in the work was a SARIMA $(0,1,1)\times(1,0,0)_{12}$. It can concluded that the EWMA chart is substantially faster than \bar{X} and R graphics to detect process disturbances allowing to complement epidemiological investigations.

Key-words: epidemiological surveillance, endemic disease, statistical control charts, SARIMA model.

LISTA DE TABELAS

Tabela 1 – Casos autóctones de hepatite C, segundo o ano e o mês do Estado do Rio Grande do Sul. Brasil, 2007 a 2011.....	66
Tabela 2 – Principais resultados do modelo SARIMA $(2,1,0) \times (1,0,0)_{12}$ ajustado para número de observações mensais de hepatite C no Rio Grande do Sul.....	69

LISTA DE QUADROS

Quadro 1 – Regras adicionais para os testes de controle.....	24
--	----

LISTA DE FIGURAS

Figura 1 – Limites três-sigma para o gráfico \bar{x}	22
Figura 2 – Diagrama de controle para incidência de poliomielite, Estado da Califórnia, EUA, 1943.....	58
Figura 3 – Gráfico da série de casos mensais de Hepatite C do Rio Grande do Sul, nos anos de 2007 a 2011.....	67
Figura 4 – Função de autocorrelação dos casos mensais de hepatite C, Rio Grande do Sul, 2007 a 2011.....	68
Figura 5 – Função de autocorrelação parcial dos casos mensais de hepatite C, Rio Grande do Sul, 2007 a 2011.....	68
Figura 6 – Série histórica original e diferenciada dos casos de hepatite C, Rio Grande do sul, 2007 a 2011.....	69
Figura 7 – Gráfico de normalidade dos resíduos para o modelo SARIMA (0,1,1)×(1,0,0) ₁₂	70
Figura 8 – Função de autocorrelação do resíduos para o modelo SARIMA (0,1,1)×(1,0,0) ₁₂	70
Figura 9 – Gráfico de controle aplicado aos resíduos do modelo para o número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.....	72
Figura 10 – Gráfico de controle com parâmetros K=0,5 e H=5 aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.....	73
Figura 11 – Gráfico de controle com parâmetros L=2 e $\lambda=0,10$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.....	74
Figura 12 – Gráfico de controle com parâmetros L=2 e $\lambda=0,40$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.....	75
Figura 13 – Gráfico de controle com parâmetros L=2 e $\lambda=0,90$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.....	76

LISTA DE ABREVIATURAS E SIGLAS

AIC – Critério de Informação de Akaike
AR – Autorregressivo
ARIMA – Autorregressivo Integrado de Média Móvel
ARIMA – Modelo Autorregressivo Integrado de Médias Móveis
ARL – Average Run Length (número de amostras retiradas até a ocorrência de um alarme)
ARL_{EC} – Número de amostras retiradas até a ocorrência de um alarme falso
ARL_{FC} – Número de amostras observadas até a detecção de um verdadeiro deslocamento
ARMA – Modelo Autorregressivo de Médias Móveis
BIC – Critério de Informação Bayesiano
CEP – Controle Estatístico de Processo
C_P – Índice de Capacidade Potencial
C_{PK} – Índice de Performance
CQ – Controle de Qualidade
CUSUM – Somas Cumulativas
EWMA – Média Móvel Exponencialmente Ponderada
FAC – Função de Autocorrelação
FACP – Função de Autocorrelação Parcial
FSESP – Fundação de Serviço de Saúde Pública
Gráfico \bar{x} – Gráfico de controle para média
Gráfico c – Gráficos de defeitos por amostra
Gráfico CUSUM – Gráfico de somas cumulativas
Gráfico EWMA – Gráfico de média móvel exponencialmente ponderada
Gráfico MR – Gráfico de controle para Médias Móveis
Gráfico np – Gráfico de peças defeituosas
Gráfico p – Gráfico de fração defeituosa na amostra
Gráfico R – Gráfico de controle para amplitude
Gráfico S – Gráfico de controle para desvio padrão
Gráfico x – Gráfico de controle para medidas individuais
LC – Linha Central
LIC – Limite Inferior de Controle
LSC – Limite Superior de Controle
MA – Média Móvel
SIA/SUS – Sistema de Informações Ambulatoriais do SUS
SIH/SUS – Sistema de Informações Hospitalares do SUS
SIM – Sistema de Informações Sobre Mortalidade
SINAN – Sistema de Informação de Agravos e Notificação
SINASC – Sistema de Informações Sobre Nascidos Vivos
SUS – Sistema Único de Saúde
TAV – Taxa de Alarmes Verdadeiros

UTI – Unidade de Tratamento Intensivo
WCUSUM – Somas Cumulativas Ponderadas

LISTA DE ANEXOS

Anexo A - Fatores para o cálculo dos limites de controle (3σ) para o gráfico de médias.....	85
Anexo B - Fatores para o cálculo dos limites de controle (3σ) para o gráfico do desvio padrão.....	86
Anexo C - Fatores para o cálculo dos limites de controle (3σ) para o gráfico da amplitude.....	87
Anexo D - Modelos selecionados para os casos de Hepatite C, tamanho amostral válido, média, valores mínimo e máximo, desvio padrão, variância e valores dos critérios AIC e BIC.....	88
Anexo E - Somas acumulativas dos dados de hepatite C, C_i^+ e C_i^- . Rio Grande do Sul, 2007 a 2011.....	89
Anexo F - Lista Nacional de Doenças e Agravos de Notificação Compulsória, Secretaria de Vigilância em Saúde /MS, 2002.....	91

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Objetivos	16
1.1.1 Objetivo geral	16
1.2.2 Objetivos específicos.....	16
1.2 Justificativa	16
1.3 Estrutura do trabalho	17
2 REVISÃO DA LITERATURA	18
2.1 Gráficos de Controle	18
2.1.1 Qualidade	18
2.1.2 Processo	18
2.1.3 Variabilidade do Processo.....	19
2.1.4 Teorema do Limite Central	20
2.1.5 Controle Estatístico de Processo	21
2.1.6 Gráficos estatísticos de controle	21
2.1.7 Interpretação dos gráficos de controle	23
2.1.8 Tipos de gráficos de controle	24
2.1.9 Gráficos de controle para variáveis	25
2.1.10 Gráficos de controle para \bar{x} e R	26
2.1.11 Gráficos de controle para \bar{x} e S	29
2.1.12 Gráfico do valor individual e amplitude móvel (x e MR)	31
2.1.13 Gráfico de Controle das Somas Cumulativas (CUSUM)	33
2.1.14 Gráfico de Controle CUSUM tabular	34
2.1.15 Gráfico de controle de EWMA	36
2.1.16 Gráficos de controle de Atributos	38
2.1.17 Gráficos de controle de Resíduos	41
2.1.18 Capacidade do Processo	42
2.1.19 Índice de capacidade potencial (C_p)	42
2.1.20 Índice de performance (C_{pk})	43
2.1.21 ARL em controle e o ARL fora de controle	44
2.1.22 Função de autocorrelação.....	45
2.1.23 Modelagem de séries temporais de acordo com Box e Jenkins.....	47
2.1.24 Processo estacionário	48
2.1.25 Autocorrelação	48
2.1.26 Modelos Univariados	49
2.1.27 Etapas da metodologia Box-Jenkins	53
2.2 Vigilância Epidemiológica	54
2.2.1 Prevalência.....	56
2.2.2 Incidência	56
2.2.3 Mortalidade.....	57
2.2.4 Uso dos Gráficos de controle em estudos Epidemiológicos.....	57
3 METODOLOGIA	64
3.1 Fontes de dados	64
3.2 Caracterização e limitação da pesquisa	64
3.3 Estratégia Analítica	65
4 RESULTADOS	66
4.1 Análise Descritiva	66

4.2 Ajuste do modelo Box e Jenkins	67
4.3 Aplicação dos Gráficos de Controle	71
4.3.1 Gráfico para medidas individuais	71
4.3.2 Gráfico de CUSUM.....	72
4.3.3 Gráfico de EWMA.....	73
5 CONCLUSÃO E CONSIDERAÇÕES FINAIS	77
REFERÊNCIAS	79
ANEXOS	85

1 INTRODUÇÃO

O Controle Estatístico de Processo (CEP) é um método quantitativo para monitorar um processo repetitivo a fim de determinar se um processo está operando adequadamente. O CEP utiliza coletas de dados de processo em tempo real e compara as medições atuais com os medidores básicos de desempenho do processo (DAVIS; AQUILINO; CHASE, 2001).

Nesse método, a média, a amplitude e o desvio padrão são as estatísticas usadas com maior frequência para analisar os dados medidos, e os gráficos de controle são utilizados para monitorar essas estatísticas. Um ponto fora do controle para qualquer uma dessas estatísticas é a indicação da presença de uma causa especial de variação e, portanto, da necessidade de uma investigação imediata a fim de identificá-la (PYZDEK; KELLER, 2011).

Os gráficos estatísticos de controle de processo foram inicialmente desenvolvidos por Walter A. Shewhart, em 1924, para serem usados em processos químicos e industriais. A partir disso, novos tipos de gráficos continuam a ser desenvolvidos como por exemplo o Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) e o Gráfico de Controle das Somas Acumuladas (CUSUM), além dos gráficos multivariados.

Entretanto, ao longo dos anos, esses gráficos têm sido teoricamente aprimorados e aplicados em muitas situações, inclusive na área da saúde.

A utilização dos Diagramas de Controle para o acompanhamento de doenças é justificada pela premissa de que a produção das doenças pode ser considerada análoga à produção do processo industrial. Ao utilizar uma ferramenta da escala industrial na análise e acompanhamento do processo biológico, supõe-se que algumas características desses dois processos sejam replicadas. Dentre essas características estão a variabilidade das unidades amostrais, que no processo biológico é representada pela variabilidade do processo de produção da doença, e a utilização de limites máximo e mínimo de controle (ALVES, 2004).

Assim, essa ferramenta estatística se torna relevante no que se refere à saúde humana, pois permite detectar possíveis alterações bruscas nos valores dos indicadores analisados e, quando observado um padrão de anormalidade, pode-se

decidir se as mesmas ocorrem por mero acaso ou não, neste último caso, apresenta subsídios para uma possível intervenção em saúde pública.

1.1 Objetivos

1.1.1 Objetivo geral

Realizar uma revisão dos princípios teóricos dos gráficos de controle estatístico de modo a serem úteis na análise de dados de vigilância epidemiológica em saúde pública.

1.2.2 Objetivos específicos

- Revisar o referencial teórico acerca de alguns gráficos de controle que podem ser utilizados na área da saúde: gráficos para atributos e gráficos para variáveis como Shewhart (para medidas individuais), Gráfico das Somas Acumuladas tabular (CUSUM tabular) e Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA);
- Identificar os principais gráficos de controle estatístico propostos na literatura utilizados em serviços de vigilância de saúde;
- Apresentar uma aplicação prática desses procedimentos com a finalidade de exemplificar o uso de gráficos de controle apresentados comparando seus resultados.

1.2 Justificativa

Os diagramas de controle, quando implementados de forma correta, constituem-se em importantes ferramentas na tomada de decisão em tempo real,

pois através deles é possível identificar tendências que sinalizam o surgimento de novos problemas e também identificar epidemias levando à ações de controle e prevenção.

O tema proposto tem sua relevância fundamentada no fato de que os gráficos abordados são procedimentos de controle estatístico eficazes e simples de serem utilizados e também por ser um campo de trabalho subexplorado, frente ao perfil epidemiológico brasileiro, no contexto dos procedimentos estatísticos aplicados à análise de dados em vigilância em saúde pública.

1.3 Estrutura do trabalho

Esta monografia está estruturada em cinco capítulos. O primeiro consta da introdução, objetivos geral e específico, justificativa e estrutura da pesquisa.

O segundo capítulo é subdividido em duas partes específicas onde a primeira consta da fundamentação teórica dos principais gráficos de controle utilizados no Controle Estatístico de Processo e a segunda parte abrange uma revisão de trabalhos sobre gráficos de controle aplicados aos dados de vigilância epidemiológica presentes na literatura.

O terceiro capítulo abrange a metodologia utilizada no presente trabalho.

No quarto capítulo se encontra uma aplicação desta metodologia usando dados de incidência de Hepatite C no Rio Grande do Sul no período de 2007 a 2011 bem como os resultados do uso de três diferentes gráficos de controle.

O último capítulo é composto pela conclusão deste trabalho e sugestões para futuras pesquisas.

2 REVISÃO DA LITERATURA

2.1 Gráficos de Controle

2.1.1 Qualidade

Por muitos anos após a II Guerra Mundial, a qualidade foi vista mais como uma função defensiva do que como uma arma competitiva para utilização no desenvolvimento de novos mercados ou no aumento da participação nos mercados já conquistados.

Dessa forma, a ênfase da qualidade estava no Controle da Qualidade (CQ) o qual era associado à redução do número de reclamações dos clientes. Como resultado, era dada maior importância à inspeção (separar produtos defeituosos dos não-defeituosos) do que a prevenção.

A identificação desses produtos e seu tratamento (retrabalho) ou eliminação (refugo) acarretavam em custos. Conseqüentemente, acreditava-se que adquirir uma qualidade melhor seria mais dispendioso. Muitas vezes, as medições de especificações do produto eram realizadas apenas nas saídas, deixando os gerentes de produção com pouco ou nenhum poder para interromper a produção ou para evitar a expedição de produtos defeituosos (DAVIS; AQUILINO; CHASE, 2001).

Para Montgomery (2004), a qualidade é uma entidade multifacetada e explicar o que é qualidade não é uma questão simples, mas o autor define que qualidade significa adequação para o uso e, em uma definição mais moderna, aponta que qualidade é inversamente proporcional à variabilidade.

2.1.2 Processo

Um processo é definido como qualquer passo ou conjunto de passos que estão envolvidos na conversão ou na transformação de insumos em resultados (DAVIS; AQUILINO; CHASE, 2001). O processo pode ainda ser entendido como um conjunto de atividades inter-relacionadas e interagentes que transformam entradas em saídas. Nesta perspectiva, a organização é vista como uma rede de processos que devem ser continuamente melhorados de forma a aumentar o valor acrescentado para todas as partes interessadas no seu desempenho (PEREIRA; REQUEIJO, 2008).

2.1.3 Variabilidade do Processo

Todo o processo varia de forma distinta. Alguns processos variam amplamente, alguns variam de maneira tênue. Uma cuidadosa análise da variação inerente em um dado processo torna possível a comparação de seu desempenho atual com seu desempenho esperado, determinado a partir do desenvolvimento passado do processo. (DAVIS; AQUILINO; CHASE, 2001).

Dessa forma, por meio do uso de experiências do passado, podemos prever, pelo menos dentro dos limites, como o fenômeno poderá variar no futuro. Assim, fica entendido que a previsão dentro de limites significa que pode-se afirmar, pelo menos de forma aproximada, a probabilidade com que o fenômeno observado irá cair dentro dos limites dados (SHEWHART, 1931).

O grande mérito de Shewhart, foi perceber que qualquer processo produtivo, mesmo que esforços específicos sejam direcionados para controlar variabilidade dos fatores existentes, sempre existirá a variabilidade natural do processo, por mais bem projetado e calculado que este seja (MONTGOMERY, 1997; COSTA, EPPRECHT, CARPINETTI, 2004).

Esta variabilidade é causada pelas variáveis não controláveis do processo, sendo comumente chamadas de causas comuns ou aleatórias e que não podem ser facilmente removidas do processo. Diz-se que um processo que opera apenas com as causas aleatórias da variação está sob controle estatístico.

Existem também as causas especiais (ou assinaláveis) de variação as quais surgem esporadicamente devido a uma situação particular que faz com que o

processo se comporte de um modo completamente diferente do usual, o que pode resultar em um deslocamento do seu nível de qualidade. Diz-se que um processo que opera na presença de causas atribuíveis está fora de controle (WERKEMA, 1995).

Verifica-se, portanto, que o controle não é definido como a completa ausência de variação e sim um estado onde toda a variação é previsível.

2.1.4 Teorema do Limite Central

Uma das distribuições mais importantes em Controle Estatístico de Processo é a Distribuição Amostral de Médias. Se X for uma variável aleatória Normalmente distribuída com média μ e variância σ^2 , a distribuição de médias de amostras de dimensão n será também Normal com média igual μ e variância $\sigma_{\bar{x}}^2 = \sigma^2/n$.

Geralmente $\sigma_{\bar{x}}^2 = \sigma^2/n$ é designado por erro padrão. Nos casos em que σ (desvio padrão da população) é desconhecido, este pode ser estimado a partir do desvio padrão amostral. Em Controle Estatístico de Processo assume-se geralmente que a distribuição de médias é Normal (PEREIRA; REQUEIJO, 2008). Este pressuposto baseia-se no chamado Teorema Central do Limite o qual é definido por Montgomery (1997) da seguinte forma:

Se X_1, X_2, \dots, X_n são variáveis aleatórias independentes com média μ_i e variância σ_i^2 e se $y = x_1 + x_2 + \dots + x_n$, então a distribuição de

$$\frac{y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad (1)$$

se aproxima da distribuição $N(0,1)$ à medida que n tende a infinito.

Em outras palavras, o teorema central do limite afirma que a soma de n variáveis aleatórias independentes é aproximadamente normal, desde que n seja suficientemente grande

Ou ainda pode-se entender que independente do formato da população ou universo, a distribuição dos valores médios das amostras coletadas daquele universo tenderá a uma distribuição normal à medida que o tamanho da amostra cresce de forma ilimitada (PYZDEK; KELLER, 2011).

2.1.5 Controle Estatístico de Processo

O Controle Estatístico de Processo (CEP) é formado por uma série de ferramentas de resolução de problemas útil na obtenção da estabilidade do processo e na melhoria da sua capacidade. O CEP pode ser aplicado a qualquer processo e dentre as suas principais ferramentas, o gráfico de controle é, provavelmente, a mais sofisticada tecnicamente (MONTGOMERY, 2004).

Através dos gráficos de controle estatístico, é possível monitorar o comportamento do processo com a finalidade de reduzir a variabilidade e determinar, a partir de estimativas dos parâmetros do processo, se este é capaz de produzir de acordo com especificações pré-definidas (PEREIRA; REQUEIJO, 2008).

2.1.6 Gráficos estatísticos de controle

Gráficos (cartas ou diagramas) de controle são instrumentos que mostram a evolução do nível de operação do processo e sua variação ao longo do tempo. Estes gráficos fornecerão estimativas da situação do processo (ou seja, sua média ou mediana) e sua variação (desvio padrão do processo).

Os gráficos de CEP são uma categoria especializada de gráficos de controle que estão baseados no teorema do limite central. O gráfico de controle típico exibe três linhas paralelas: a central, que representa o valor médio da característica de qualidade correspondente à situação do processo sob controle, isto é, sob a atuação de apenas causas de variação aleatórias; a superior, que representa o Limite Superior de Controle (LSC); e a inferior, que representa o Limite Inferior de Controle (LIC).

O valor médio para o processo é calculado (muitas vezes com base em desempenhos passados) e, dependendo do tamanho de amostra que estiver sendo utilizado, são calculados os limites de controle.

Estes limites de controle são geralmente estabelecidos em três desvios-padrão (3σ) acima e abaixo da média, pois aproximadamente três desvios-padrão do valor médio representam 99,7% da área, sob uma distribuição normal. Segundo

Montgomery (2004), os limites são escolhidos de modo que, se o processo está sob controle, praticamente todos os pontos amostrais estarão entre eles (Figura 1).

Os pontos plotados no gráfico representam amostras tomadas em momentos diferentes. No eixo x do gráfico de CEP, encontra-se o tempo; no eixo y , está a variação a partir do valor médio do processo. É usual unir os pontos por segmentos de reta, para melhor visualizar a evolução da característica de qualidade ao longo do tempo.

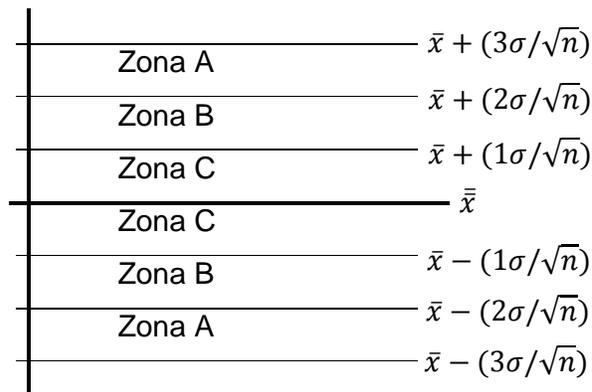


Figura 1 - Limites três-sigma para o gráfico \bar{x} .
Fonte: Montgomery,2004.

Segundo Alwan e Roberts (1988), duas suposições precisam ser satisfeitas, independentemente do tipo do gráfico de controle utilizado, para que os resultados sejam válidos:

a) É necessário que as observações sejam independentes e identicamente distribuídas, ou seja, que as amostras sejam retiradas de forma aleatória e que o processo, que as gerou, esteja sob controle estatístico;

b) As observações devem seguir alguma distribuição de probabilidade específica, tais como Normal, Binomial ou Poisson.

Segundo os autores, essas pressuposições, que são a base de toda a teoria desses gráficos, tornaram os gráficos de controle ferramentas simples porém poderosas.

2.1.7 Interpretação dos gráficos de controle

Os gráficos de controle são usados para verificar se um processo está estável, ou se uma causa especial está presente no processo (processo instável ou fora de controle). Dessa forma, o gráfico de controle estatístico de processo deve ser analisado com a intenção de evidenciar um comportamento não-aleatório (ou sistemático) indicando a existência de causas especiais de variação.

Inicialmente, um processo está fora de controle quando um ponto se encontra acima ou abaixo do limite de controle inferior. Entretanto, ao longo dos anos têm sido publicadas novas regras que permitem reconhecer padrões sistematizados e que se baseiam na probabilidade (mínima) de ocorrência de cada um dos padrões, são as chamadas regras adicionais.

Essas regras estão amplamente presentes na literatura (JURAN, 1992; WERKEMA, 1995; DAVIS, AQUILINO, CHASE, 2001; MONTGOMERY, 2004; PEREIRA e REQUEIJO, 2008) porém as várias regras publicadas não diferem substancialmente entre si.

Para a aplicação destas regras, considera-se que o intervalo compreendido entre os limites de controle está dividido em seis zonas iguais, cada uma com uma largura equivalente a 1σ e designadas por A, B, C (onde a zona C está a um desvio padrão da média, a zona B está entre um e dois desvios padrão da média e a Zona A está entre dois e três desvios padrão da média). O Quadro 1 ilustra as regras adicionais para detecção de causas especiais.

Um processo que apresenta variação não-aleatória é dito como sendo fora do controle, porém não significa que esteja sendo produzidos produtos ruins. Quando é verificada a evidência de variação não-aleatória em um processo deve-se identificar os fatores que podem ter causado esta variação, uma vez que essas regras destinam-se, essencialmente, a aumentar a sensibilidade dos gráficos de Shewhart para detectar essas causas especiais.

1. Um ponto qualquer fora dos limites de ação (limites $\pm 3\sigma$)
2. Nove pontos sucessivos em um mesmo lado da linha central
3. Seis pontos consecutivos no sentido ascendente ou descendente
4. Quatorze pontos em sequência crescendo e decrescendo alternadamente
5. Dois entre três pontos sucessivos em um mesmo lado da linha central na zona A ou além
6. Quatro entre cinco pontos sucessivos em um mesmo lado da linha central na zona B ou além
7. Quinze pontos em desordem dentro da zona C (acima e abaixo da linha central)
8. Oito pontos consecutivos de ambos os lados da linha central, sem nenhum na zona C

Quadro 1 - Regras adicionais para os testes de controle.
Fonte: Pereira e Requeijo, 2008.

Para Pereira e Requeijo (2008), estas regras são úteis em uma fase inicial de implementação do CEP, quando o processo não está estabilizado, mas o uso contínuo e frequente das regras para aumentar a sensibilidade dos gráficos, quando se deseja detectar alterações pequenas, não é aconselhável. Nestes casos, o Gráfico das Somas Acumuladas (CUSUM) e o Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) são mais adequados.

2.1.8 Tipos de gráficos de controle

Os gráficos de controle podem ser classificados conforme o tipo de característica de qualidade a ser monitorada. Basicamente existem dois tipos de gráficos de controle frequentemente utilizados que, na terminologia de Controle Estatístico de Processo, designam-se por gráficos de controle para variáveis e gráficos de controle de atributos. Ainda de acordo com a mesma terminologia, variáveis e atributos são definidos da seguinte forma:

- Variáveis: são todas as características susceptíveis de serem expressas numa escala contínua (dimensões, peso); neste caso, dado que a dispersão da população não depende unicamente da medida de tendência central, devem ser

contruídos dois gráficos de controle, um para controlar o parâmetro de localização e outro para controlar o parâmetro de dispersão da população.

- Atributos: são características que, não podendo ser expressas numa escala contínua, assumem apenas valores discretos, como o número de unidades não conformes (0, 1, 2, ...) detectadas numa linha de produção ou o número de defeitos encontrados numa ou mais unidades do produto; e evidente que devem ser definidos, caso a caso, os critérios que permitem classificar uma unidade como não conforme, bem como especificar o que se considera um defeito. Neste caso é construído apenas um gráfico de controle pois o desvio padrão está em função da média e da dimensão da amostra.

Os valores das estatísticas a serem plotados no gráfico de controle são obtidos a partir de cada uma das amostras coletadas do processo em períodos de tempo determinado pelos subgrupos racionais.

O conceito de subgrupos racionais refere-se aos subgrupos ou amostras que devem ser selecionados de forma que, se neles presentes causas especiais, a chance de diferenças entre subgrupos será maximizada, enquanto a chance de diferenças em detrimento das causas especiais dentro de um subgrupo será minimizada. Em muitas situações, o subgrupo racional consistirá em uma única observação.

2.1.9 Gráficos de controle para variáveis

Os gráficos de controle de variáveis devem ser usados sempre que seja economicamente viável o controle de características medidas em uma escala contínua. Segundo Pereira e Requeijo (2008) estes gráficos são bastante úteis por vários motivos:

- Há muitos processos e produtos com características que são mensuráveis numa escala contínua;

- Uma medição contém muito mais informação do que uma avaliação feita em termos de sim-não, passa-não passa, conforme-não conforme.

- Embora seja mais caro recolher informação sobre variáveis contínuas, é necessário, em geral, recolher menos dados e, por vezes, o custo total é menor do que os gráficos de controle de atributos;

- Devido ao fato de se necessitar de menos informação, o tempo para detecção de um problema e a sua resolução é também menor, dando origem a menos produtos não conforme;

- Mesmo que todas as observações individuais estejam dentro das especificações, o comportamento global do processo pode ser analisado, o que é muito importante na busca de soluções que melhorem continuamente o desempenho do processo e a qualidade do produto.

A seguir são apresentados os gráficos de controle para variáveis, com três desvios padrões de afastamento em relação à linha média.

2.1.10 Gráficos de controle para \bar{x} e R

O gráfico \bar{x} é usado para monitorar processos onde a média amostral é a estatística de teste. Este gráfico é também chamado de gráfico de média e tem o propósito de monitorar o nível do processo. Supõe-se que a característica de qualidade tenha uma distribuição normal com média μ e desvio padrão σ . Na prática, em geral, os parâmetros μ e σ não são conhecidos, então deve ser utilizadas estimativas disponíveis \bar{x} e S . Estas estatísticas são expressas da seguinte forma:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2a, 2b)$$

E como a estatística z , tem distribuição normal padrão $N(0,1)$ onde:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (3)$$

Tem-se que

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad (4)$$

que é o intervalo de confiança de nível $1 - \alpha$ para o parâmetro μ (média do processo). Sabe-se que \bar{x} é normalmente distribuída, com média μ e desvio padrão $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Assim, pode-se utilizar os limites do intervalo de confiança, como limites do gráfico de controle \bar{x} , e isto é:

$$LSC = \mu_{\bar{x}} + 3\sigma_{\bar{x}} = \bar{x} + 3\frac{\sigma}{\sqrt{n}} \quad (5a)$$

$$LC = \mu_{\bar{x}} = \bar{x} \quad (5b)$$

$$LIC = \mu_{\bar{x}} - 3\sigma_{\bar{x}} = \bar{x} - 3\frac{\sigma}{\sqrt{n}} \quad (5c)$$

Na prática, em geral não conhecemos μ e σ . Então, eles devem ser estimados a partir de amostras retiradas quando o processo supostamente estava sob controle. Suponha que m amostras estejam disponíveis, cada uma com n observações da característica de qualidade. Tais estimativas devem se basear em pelo menos 20 ou 25 amostras (MONTGOMERY, 2004). Tipicamente n será pequeno, frequentemente 4, 5 ou 6. Esses pequenos tamanhos de amostra resultam em pequenos subgrupos denominados subgrupos racionais. Sejam $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ as médias de cada uma das amostras. Então, o melhor estimador de μ , a média do processo, é a média geral, isto é:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (6)$$

Assim, $\bar{\bar{x}}$ deve ser usado como a linha central no gráfico \bar{x} .

O estimador do desvio padrão σ pode ser obtido através das amplitudes das m amostras ou através dos desvios padrão. Se x_1, x_2, \dots, x_n é uma amostra de tamanho n , então a amplitude da amostra é a diferença entre a maior e a menor observação como segue:

$$R = x_{m\acute{a}x} - x_{m\acute{i}n} \quad (7)$$

Sejam R_1, R_2, \dots, R_m as amplitudes das m amostras. A amplitude média é:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (8)$$

Existe uma relação bem conhecida entre a amplitude de uma amostra de uma distribuição normal e o desvio padrão de tal distribuição chamada de amplitude relativa que é dada por $W = \bar{R}/d_2$. Os parâmetros da distribuição de W são funções

do tamanho da amostra n . A média de W é uma constante d_2 , onde os valores de d_2 são tabelados (ver Anexo C). Consequentemente, um estimador de σ é $\hat{\sigma} = R/d_2$.

Assim, se \bar{R} é a amplitude média das m amostras preliminares, σ pode ser estimado usando $\hat{\sigma} = \bar{R}/d_2$, que é um estimador não viesado de σ .

Usando $\bar{\bar{x}}$ como estimador de μ e \bar{R}/d_2 como estimador de σ , então os parâmetros do gráfico \bar{x} são:

$$LSC = \bar{\bar{x}} + \frac{3}{d_2\sqrt{n}}\bar{R} \quad (9a)$$

$$LC = \bar{\bar{x}} \quad (9b)$$

$$LSI = \bar{\bar{x}} - \frac{3}{d_2\sqrt{n}}\bar{R} \quad (9c)$$

Para facilitar, faz-se $A_2 = \frac{3}{d_2\sqrt{n}}$, cujos valores também são tabelados em função do tamanho n da amostra (ver Anexo A). Então os limites de controle se reduzem à:

$$LSC = \bar{\bar{x}} + A_2\bar{R} \quad (10a)$$

$$LC = \bar{\bar{x}} \quad (10b)$$

$$LSI = \bar{\bar{x}} - A_2\bar{R} \quad (10c)$$

Considerando agora o gráfico R , a linha central será \bar{R} . Para determinar os limites de controle, é necessário uma estimativa de σ_R que pode ser determinado a partir da distribuição da amplitude relativa $W = R/\sigma$, supondo que a característica controlada seja normalmente distribuída. O desvio padrão de W , denotado por d_3 , é uma função conhecida de n , então, como $R = W\sigma$, o desvio padrão de R é:

$$\sigma_R = d_3\sigma \quad (11)$$

Como σ é conhecido, deve-se estimar σ_R por:

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \quad (12)$$

Consequentemente, os parâmetros do gráfico R com intervalos de controle contendo 3-sigma são:

$$LSC = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3d_3 \frac{\bar{R}}{d_2} \quad (13a)$$

$$LC = \bar{R} \quad (13b)$$

$$LSI = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3 \frac{\bar{R}}{d_2} \quad (13c)$$

Definindo os seguintes coeficientes:

$$D_3 = 1 - 3 \frac{d_3}{d_2} \quad (14)$$

$$D_4 = 1 + 3 \frac{d_3}{d_2} \quad (15)$$

Então os limites de controle para o Gráfico R se reduzem à:

$$LSC = D_4 \bar{R} \quad (16a)$$

$$LC = \bar{R} \quad (16b)$$

$$LIC = D_3 \bar{R} \quad (16c)$$

2.1.11 Gráficos de controle para \bar{x} e S

Similarmente ao gráfico da amplitude, o gráfico do desvio padrão S em conjunto com o gráfico \bar{x} também tem a função de detectar alterações na dispersão do processo. Algumas vezes torna-se desejável estimar diretamente o desvio padrão do processo em vez de inderamente através do uso da amplitude R . A média (\bar{x}) e a variância (s^2) amostrais são estimadores não-viesados da média e da variância populacionais (μ e σ^2). Isto é,

$$E(\bar{x}) = \mu \quad \text{e} \quad E(S^2) = \sigma^2 \quad (17a \text{ e } b)$$

onde o operador E é simplesmente o operador do valor esperado (modo abreviado de denotar o processo de calcular a média de uma variável aleatória).

Já o desvio padrão

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (18)$$

é considerado um estimador tendencioso do desvio padrão populacional σ . A distribuição amostral do desvio padrão S segue uma qui-quadrado com $v = n - 1$

graus de liberdade, mas adotando-se o intervalo de 3-sigma da distribuição normal, para faixa de controle, verifica-se que a aproximação normal é satisfatória, podendo ser tabelados coeficientes em função do tamanho da amostra n :

$$E(S) = \left(\frac{2}{n-1}\right)^{\frac{1}{2}} \frac{\Gamma(n/2)}{\Gamma[(n-1/2)]} \sigma \quad (19)$$

Com $c_4 = \frac{\Gamma(n/2)}{\Gamma[(n-1/2)]}$ tem-se $E(S) = c_4 \sigma$. Onde c_4 é um valor tabelado e os valores desta constante encontram-se listadas no Anexo B.

Caso não exista nenhum valor de referência para σ , é necessário estimá-lo através de dados passados. Suponha que m amostras preliminares estejam disponíveis, cada uma de tamanho n , e seja S_i o desvio padrão da i^{a} amostra. A média dos m desvios padrões é:

$$\bar{S} = \frac{S_1 + S_2 + \dots + S_m}{m} \quad (20)$$

A estatística \bar{S}/c_4 é um estimador não-viesado de σ , e um estimador não tendencioso de $E(S)$ é $\hat{\sigma}_S = \frac{\bar{S}}{c_4}$ e um estimador não tendencioso para $\hat{\sigma}_S = \sigma \sqrt{1 - c_4^2}$ é dado por $\hat{\sigma}_S = \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2}$. Então, os parâmetros para o gráfico S são:

$$LSC = \bar{S} + 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \quad (21a)$$

$$LC = \bar{S} \quad (21b)$$

$$LIC = \bar{S} - 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \quad (21c)$$

Para facilitar as aplicações, é usual definir as constantes:

$$B_3 = 1 - 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \quad (22)$$

$$B_4 = 1 + 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \quad (23)$$

Assim, desconhecendo-se σ os limites de controle para o gráfico S são:

$$LSC = B_4 \bar{S} \quad (24a)$$

$$LC = \bar{S} \quad (24b)$$

$$LIC = B_3 \bar{S} \quad (24c)$$

Quando o estimador \bar{S}/c_4 é usado para estimar σ , os limites para o gráfico \bar{x} correspondente pode ser definido da seguinte forma:

$$LSC = \bar{\bar{x}} + \frac{3\bar{S}}{c_4\sqrt{n}} \quad (25a)$$

$$LC = \bar{\bar{x}} \quad (25b)$$

$$LIC = \bar{\bar{x}} - \frac{3\bar{S}}{c_4\sqrt{n}} \quad (25c)$$

Definindo a constante $A_3 = 3/(c_4\sqrt{n})$, os parâmetros do gráfico \bar{x} se tornam:

$$LSC = \bar{\bar{x}} + A_3\bar{S} \quad (26a)$$

$$LC = \bar{\bar{x}} \quad (26b)$$

$$LIC = \bar{\bar{x}} - A_3\bar{S} \quad (26c)$$

As constantes A_3 , B_3 e B_4 para a construção dos gráficos \bar{x} e S , obtidas com base em dados passados, são tabelados e encontram-se nos Anexos A e B. Essas constantes para cálculo dos limites de controle (A_2, D_3, D_4, \dots) são obtidas admitindo-se que os valores individuais sejam estatisticamente independentes normalmente distribuídas.

2.1.12 Gráfico do valor individual e amplitude móvel (x e MR)

Existem casos em que a amostra é formada por uma única observação individual, ou seja, $n = 1$, usa-se nestas situações, o gráfico de controle de observações individuais e amplitudes móveis (*moving range*). Estas cartas são úteis quando as medições são muito caras, quando os testes são destrutivos, quando não é possível o recolhimento de amostras com mais de uma observação (caso de muitos processos químicos) ou ainda quando há sistemas computadorizados de medição e são feitas medições de todas as unidades.

Com observações individuais não é possível calcular uma estatística que dê indicações sobre a dispersão do processo. Assim, calcula-se a amplitude móvel de duas observações consecutivas como base para estimar a variabilidade do processo. A amplitude móvel é definida como:

$$MR_i = |x_i - x_{i-1}| \quad (27)$$

A linha central do gráfico X é a média dos m valores da variável

$$LC_x = \bar{x} = \frac{1}{m} \sum_{i=1}^m X_i \quad (28)$$

e a linha central do gráfico MR é a média das $(m - 1)$ amplitudes móveis

$$LC_{MR} = \overline{MR} = \frac{1}{m-1} \sum_{i=1}^{m-1} MR_i \quad (29)$$

Quando os parâmetros do processo não são conhecidos, utiliza-se \overline{MR}/d_2 como estimador de σ , à semelhança do que foi feito para as cartas \bar{x} e R . Para o gráfico de controle de observações individuais (x), os parâmetros são

$$LSC_x = \bar{x} + 3\sigma_x = \bar{x} + \frac{3\overline{MR}}{d_2} \quad (30a)$$

$$LC = \bar{x} \quad (30b)$$

$$LIC_x = \bar{x} - 3\sigma_x = \bar{x} - \frac{3\overline{MR}}{d_2} \quad (30c)$$

O gráfico de Amplitudes Móveis (MR) é então definido por:

$$LSC_{MR} = D_4 \overline{MR} \quad (31a)$$

$$LC_{MR} = \overline{MR} \quad (31b)$$

$$LIC_{MR} = D_3 \overline{MR} \quad (31c)$$

Nestes gráficos, d_2 , D_3 e D_4 terão valores correspondentes a uma dimensão de amostra igual ao número de observações usadas para calcular cada amplitude móvel; por exemplo, se a amplitude móvel é formada por duas observações ($n = 2$), é usada, então $d_2 = 1,128$, $D_3 = 0$ e $D_4 = 3,267$.

Deve-se tomar um certo cuidado na interpretação de padrões no gráfico das amplitudes móveis. As amplitudes móveis são correlacionadas e essa correlação pode muitas vezes induzir um padrão de sequências ou ciclos nos gráficos de controle. Também, supõe-se que as medidas individuais no gráfico x não sejam correlacionadas e qualquer padrão aparente nessa carta deve ser cuidadosamente investigado.

2.1.13 Gráfico de Controle das Somas Cumulativas (CUSUM)

Os gráficos de controle de soma acumulada (Cumulative Sum Control Charts - CUSUM) são indicados para o monitoramento de processos sujeitos a pequenas perturbações. Proposto por Page (1954), este gráfico representa uma soma acumulada dos valores do processo em análise, de forma que todos os dados são utilizados e interferem na construção e plotagem no gráfico.

Estes gráficos guardam informações acumuladas de toda a sequência de pontos e, por isso, são mais sensíveis para detectar pequenos desvios da média de um processo. Dessa forma, uma pequena mudança que se apresente por algumas observações em sequência será percebida por sua soma, enquanto que no gráfico de Shewhart essas alterações podem não ser detectadas.

O gráfico de controle CUSUM, segundo Costa, Epprecht e Carpinetti (2004), utiliza informações de diversas amostras de tamanho $n \geq 1$ para decidir sobre o estado do processo: à medida que as amostras são retiradas, os desvios de \bar{x} em relação ao valor-alvo μ_0 (valor médio em controle ou valor nominal) são acumulados, gerando a estatística C_i , onde

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0), \quad i = 1, 2, \dots \quad (32)$$

é definida como soma cumulativa, dada pela soma acumulada das diferenças entre observações de uma série e um valor de referência previamente estabelecido (μ_0) até a i -ésima amostra, inclusive. Montgomery (2004) salienta que os gráficos CUSUM são particularmente eficientes quando as amostras são de tamanho $n = 1$.

Em um processo sob controle para o valor médio desejável μ_0 , espera-se que a soma cumulativa definida na equação anterior oscile em torno de zero. Entretanto, se o valor de C_i tender a crescer ($\mu_1 > \mu_0$) será observada uma tendência positiva na soma cumulativa. De maneira análoga, se C_i tender a diminuir ($\mu_1 < \mu_0$), apresentará uma tendência negativa.

Nesses casos onde existe uma tendência se desenvolvendo nos pontos plotados, deve-se considerar esse fato como evidência de que a média do processo mudou, e uma pesquisa para buscar as causas assinaláveis deve ser realizada.

Os gráficos de CUSUM podem ser representados de duas maneiras: o CUSUM tabular (ou algorítmico) e a forma da máscara de V do CUSUM. Das duas

representações, a tabular é a preferível, pois apesar de a máscara V ser bastante discutida na literatura, não é de fácil interpretação e apresenta outras desvantagens no seu uso (CASTILLO, 2002).

2.1.14 Gráfico de Controle CUSUM tabular

Os CUSUMs podem ser construídos tanto para observações individuais quanto para médias de subgrupos. Seja x_i a i -ésima observação do processo. Quando está sob controle, x_i tem uma distribuição normal com média μ_0 e desvio padrão σ . Assume-se que σ é conhecido ou de estimação possível.

O CUSUM tabular trabalha acumulando tanto os desvios μ_0 que estão acima do alvo, com uma estatística C^+ , e acumulando desvios em relação à μ_0 , quanto os que estão abaixo do alvo, com outra estatística C^- . As estatísticas C^+ e C^- são designadas por somas unilaterais superior e inferior, respectivamente, e são calculadas conforme as seguintes equações:

$$C_i^+ = \text{máx} [0, x_i - (\mu_0 + K) + C_{i-1}^+] \quad (33a)$$

$$C_i^- = \text{máx} [0, (\mu_0 - K) - x_i + C_{i-1}^-] \quad (33b)$$

cujos valores iniciais são $C_0^+ = C_0^- = 0$, e K é um valor de referência (valor de compensação, folga ou tolerância) e vale aproximadamente a metade do valor que se tem interesse em detectar rapidamente, dessa forma, frequentemente é escolhido o valor correspondente a metade da amplitude entre μ_0 (valor pretendido) e μ_1 (valor da média fora de controle).

Assim, se a mudança é expressa em termos do número de desvios padrão quando $\mu_1 = \mu_0$, então K , o fator de sensibilidade, representa a metade da magnitude dessa mudança, ou seja,

$$K = \frac{\delta}{2} \sigma = \frac{|\mu_1 - \mu_0|}{2} \quad (34)$$

onde δ é o tamanho da mudança que se deseja detectar em unidades de desvio padrão; σ o desvio padrão; μ_0 o valor pretendido e μ_1 o valor da média fora de controle.

O valor de H , o intervalo de decisão, é normalmente atribuído em números de desvio padrão, alguns autores (CASTILLO, 2002; MONTGOMERY, 2004; COSTA, EPPRECHT, CARPINETTI, 2004) recomendam de quatro a cinco vezes o valor do desvio padrão σ do processo controlado.

O CUSUM tabular é um procedimento que usa duas somas individuais, unilaterais C_i^+ e C_i^- , e as compara com um intervalo de decisão H . Caso uma das amostras seja maior que o valor H , o processo é dito fora de controle estatístico. Esse método também indica quando a mudança provavelmente ocorre pois é possível usar contadores (N_i^+ e N_i^-) que indicam o número de períodos em que a soma unilateral C_i^+ ou C_i^- foi diferente de zero.

O gráfico de CUSUM, algumas vezes chamados de gráfico de status do CUSUM, são construídos pela plotagem de C_i^+ e C_i^- versus o número da amostra. As barras verticais do gráfico representam os valores de C_i^+ e C_i^- , no período i . Com o intervalo de decisão plotado no gráfico, o gráfico de status de CUSUM se assemelha ao gráfico de Shewhart. As observações x_i de cada período, também podem ser plotadas no gráfico, o que facilita a visualização do desempenho do processo.

A ação a ser tomada quando um sinal de descontrole é identificado em um esquema CUSUM é semelhante à de qualquer gráfico de controle: pesquisar as causas assinaláveis, agir de forma a corrigir as falhas e então reiniciar o algoritmo das somas acumuladas.

Em situações onde se deseja o ajuste de alguma variável manipulada para trazer o processo de volta ao valor-alvo μ_0 , pode ser prático ter uma estimativa da nova média do processo após a mudança. Isso pode ser calculado por:

$$\hat{\mu} = \begin{cases} \mu_0 + K + \frac{C_i^+}{N^+} & \text{se } C_i^+ > H \\ \mu_0 - K - \frac{C_i^-}{N^-} & \text{se } C_i^- > H \end{cases} \quad (35)$$

O algoritmo CUSUM tem sua aplicação estendida aos subgrupos racionais ($n > 1$), para isso basta substituir x_i por \bar{x} e considerar σ/\sqrt{n} em vez de σ (por exemplo para estabelecer o valor de K).

2.1.15 Gráfico de controle de EWMA

O gráfico de controle da Média Móvel Exponencialmente Ponderada (EWMA) também é muito sensível em detectar pequenas e moderadas alterações dos parâmetros do processo e, portanto, é uma boa alternativa ao gráfico de controle de Shewhart. O gráfico de EWMA apresenta desempenho semelhante ao do gráfico das somas cumulativas e, como este, é geralmente usado em observações individuais sendo, de certa forma, mais fácil de se estabelecer e operar.

Introduzido por Roberts (1959), o gráfico da média móvel exponencialmente ponderada é definido como:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (36)$$

Onde $0 < \lambda \leq 1$ é uma constante e o valor inicial (exigido com a primeira amostra em $i = 1$) é o alvo do processo, de modo que $z_0 = \mu_0$. Algumas vezes, a média de dados preliminares é usada como valor inicial do EWMA, de modo que $z_0 = \bar{x}$.

Segundo Castillo (2002), quanto mais λ se aproxima de 1, mais peso é dado à média mais recente e a medida que λ diminui, mais peso é dado para os dados mais antigos. Já quando λ está no limite inferior ($\lambda = 0$), todos os z_i são iguais a z_0 .

Para demonstrar que o modelo da equação 36 constitui, de fato, uma média ponderada para todas as observações anteriores, pode-se substituir z_{i-1} no membro direito da equação, obtendo-se:

$$\begin{aligned} z_i &= \lambda x_i + (1 - \lambda)[\lambda x_{i-1} + (1 - \lambda)z_{i-2}] \\ &= \lambda x_i + \lambda(1 - \lambda)x_{i-1} + (1 - \lambda)^2 z_{i-2} \end{aligned} \quad (37)$$

Continuando a substituir recursivamente z_{i-j} , $j=2,3,\dots,t$, obtém-se:

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0 \quad (38)$$

Os pesos $\lambda(1 - \lambda)^j$ decrescem geometricamente com a idade da média amostral. Além disso, os pesos têm soma um, pois

$$\lambda \sum_{j=0}^{i-1} (1 - \lambda)^j = \lambda \left[\frac{1 - (1 - \lambda)^i}{1 - (1 - \lambda)} \right] = 1 - (1 - \lambda)^i \quad (39)$$

A Média Móvel Exponencialmente Ponderada é, algumas vezes, chamada média móvel geométrica devido a esses pesos que decrescem geometricamente quando ligados por uma curva contínua. Como a EWMA pode ser considerada como uma média móvel ponderada de todas as observações passadas e corrente, o gráfico de EWMA é insensível à hipótese de normalidade, e é o ideal para ser usado em observações individuais.

Se as observações x_i são variáveis aleatórias independentes com variância σ^2 , então a variância de z_i é dada por

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda} \right) [1 - (1 - \lambda)^{2i}] \quad (40)$$

Portanto, o gráfico de controle EWMA é construído com base nos pares ordenados (t, z_i) , verificando-se uma situação de fora de controle (alteração da média do processo) se z_i não pertencer ao intervalo definido pelos limites de controle. Esses limites são definidos por

$$LSC = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1 - \lambda)^{2i}]} \quad (41a)$$

$$LC = \mu_0 \quad (41b)$$

$$LIC = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1 - \lambda)^{2i}]} \quad (41c)$$

Onde o fator L é a largura dos limites de controle, λ é o fator de ponderação, μ_0 é um valor objetivo que pode ser substituído pela média do processo e σ é o desvio padrão do processo.

A medida que i aumenta, a quantidade $[1 - (1 - \lambda)^{2i}]$ tende para a unidade, assim os limites de controle tendem assintoticamente para

$$LSC = \mu_0 + L\sigma_0 \sqrt{\left(\frac{\lambda}{2-\lambda} \right)} \quad (42a)$$

$$LC = \mu_0 \quad (42b)$$

$$LIC = \mu_0 - L\sigma_0 \sqrt{\left(\frac{\lambda}{2-\lambda} \right)} \quad (42c)$$

Montgomery (2004) recomenda o uso dos limites de controle exatos nas equações 41a e 41c para pequenos valores de i , o que resultará em um melhor desempenho do gráfico de controle de EWMA.

Este gráfico é, em geral, usado com observações individuais. No entanto, se são tomados subgrupos racionais ($n > 1$), então deve-se substituir x_i por \bar{x}_i , e σ por $\sigma_{\bar{x}} = \sigma\sqrt{n}$ nas equações anteriores.

2.1.16 Gráficos de controle de Atributos

Existem duas situações em que se utilizam atributos, em vez de variáveis: quando não é possível medir, tais como as características inspecionadas visualmente (cor, brilho, arranhões, danos, etc.); e quando as medidas são possíveis, mas não são tomadas por questões econômicas, de tempo ou de necessidade.

Neste caso, deve-se aplicar as cartas de controle para atributos. As quatro cartas de controle por atributos, mais usadas, são segundo Montgomery (2004), as seguintes:

a) Carta de controle p: Os princípios estatísticos subjacentes à carta de controle para a fração defeituosa (ou fração não-conforme), baseiam-se na distribuição Binomial. Suponha que o processo de produção esteja operando de maneira estável, de tal modo que a probabilidade de que uma unidade não esteja de acordo com as especificações seja p , e que as sucessivas unidades produzidas sejam independentes.

Esta probabilidade é constante e os itens produzidos são variáveis aleatórias i.i.d., ou seja, independentes, identicamente distribuídas, com distribuição Bernoulli $X_i \sim b(1, p)$. O parâmetro da Bernoulli é p . Se uma amostra aleatória de n unidades do produto é selecionada, e se d é o número de unidades do produto que são não-conformes, então d tem uma distribuição binomial com parâmetros n e p , isto é

$$P(d = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, \dots, n \quad (43)$$

A fração defeituosa da amostra é definida como sendo a razão entre o número de defeituosos encontrados na amostra (d) e o tamanho da amostra (n):

$$\hat{p} = \frac{d}{n} \quad (44)$$

A distribuição de probabilidade de \hat{p} é obtida a partir da binomial. A média e a variância de \hat{p} são:

$$\mu_{\hat{p}} = p \quad (45)$$

$$\sigma_{\hat{p}} = \frac{p(1-p)}{n} \quad (46)$$

Suponha que a verdadeira fração de defeituosos, p , no processo de produção seja conhecida. Os limites de controle para três desvios padrões são $\mu(p) \pm 3\sigma p$. Então a linha central e os limites de controle da fração defeituosa são definidos como:

$$LSC = p + 3\sqrt{\frac{p(1-p)}{n}} \quad (47a)$$

$$LC = p \quad (47b)$$

$$LIC = p - 3\sqrt{\frac{p(1-p)}{n}} \quad (47c)$$

Quando a fração não-conforme, p , não é conhecida, deve-se estimá-la a partir dos dados observados. O procedimento usual é a seleção de m amostras preliminares, cada uma de tamanho n . Como regra geral, m deve ser 20 ou 25. Se há d_i unidades defeituosas na amostra i , calcula-se a fração defeituosa na i -ésima amostra como:

$$\hat{p}_i = \frac{d_i}{n} \quad i = 1, 2, \dots, m \quad (48)$$

A média das frações defeituosas das amostras individuais é:

$$\bar{p} = \frac{\sum_{i=1}^m d_i}{mn} = \frac{\sum_{i=1}^m \hat{p}_i}{m} \quad (49)$$

A estatística \bar{p} estima a fração não-conforme desconhecida, p . A linha central e os limites de controle da carta de controle para a fração defeituosa são calculados como:

$$LSC = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (50a)$$

$$LC = \bar{p} \quad (50b)$$

$$LIC = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (50c)$$

Segundo Montgomery (2004), o valor amostral de \hat{p}_i dos subgrupos preliminares devem ser plotados com limites testes para verificar se o processo estava sob controle quando foram coletados os dados preliminares. Quaisquer pontos que excedam os limites de controle tentativos devem ser investigados. Se forem descobertas causas especiais para esses pontos, eles deverão ser descartados e novos limites de controle tentativos deverão ser determinados.

b) Carta np: É similar a carta p , com a diferença que o que se deseja é marcar o número de itens defeituosos na amostra. Se não se dispõe de um valor padrão para p , então \bar{p} pode ser usado para estimar p . Como $p = d/n$, $d = n \cdot p$, logo os limites de controle são:

$$LSC = n\bar{p} + 3\sqrt{n\bar{p}(1 - \bar{p})} \quad (51a)$$

$$LC = n\bar{p} \quad (51b)$$

$$LIC = n\bar{p} - 3\sqrt{n\bar{p}(1 - \bar{p})} \quad (51c)$$

c) Carta c: Quando a característica a ser monitorada é o número total de defeitos ocorridos em uma unidade do produto, o correto é utilizar a carta de controle C . Para a operação de uma carta C , a “unidade” não necessariamente precisa ser uma unidade de determinado produto, como uma televisão ou um motor.

A unidade, denominada unidade de inspeção, pode ser um subgrupo (de tamanho constante) de itens. A construção da carta de controle para o número de defeitos segue a distribuição de Poisson. Entretanto, se $c > 5$, então, no lugar da distribuição de Poisson, pode-se utilizar a distribuição normal. Afim de que o modelo de Poisson seja aplicável, os subgrupos devem ser uniformes de forma a apresentarem probabilidades aproximadamente iguais de ocorrência de defeitos. O modelo de Poisson com parâmetro $\theta = c$ para o número de defeitos X observado por unidade do produto é:

$$P(X = k) = \frac{c^k e^{-c}}{k!}, \quad k = 0, 1, 2, \dots \quad (52)$$

A média da v.a. X é o parâmetro c e a variância também é c , logo o desvio padrão é \sqrt{c} e os limites de controle a 3 desvios padrão, são:

$$LSC = c + 3\sqrt{c} \quad (53a)$$

$$LC = c \quad (53b)$$

$$LIC = c - 3\sqrt{c} \quad (53c)$$

Como em geral o parâmetro c é desconhecido, deve-se estimá-lo com base em uma amostra preliminar de unidades de inspeção. Sendo inspecionadas n ($n \geq 20$) unidades de inspeção, uma estimativa da média e da variância de c será a média amostral \bar{c} e a carta terá então os limites de controle dados por

$$LSC = \bar{c} + 3\sqrt{\bar{c}} \quad (54a)$$

$$LC = \bar{c} \quad (54b)$$

$$LIC = \bar{c} - 3\sqrt{\bar{c}} \quad (54c)$$

Onde $\bar{c} = \frac{\sum_{i=1}^k c_i}{k}$ e k é a quantidade total de amostras e c_i o número de defeitos observado na i -ésima amostra. Quando resultar $\bar{c} - 3\sqrt{\bar{c}} < 0$, o limite inferior é colocado em zero. E ainda, quando o parâmetro c não é conhecido, os limites de controle com o estimador c , devem ser considerados como limites de controle tentativos, e as amostras preliminares examinadas em relação à falta de controle. A carta C também é conhecida como carta para o número de defeitos na amostra.

2.1.17 Gráficos de controle de Resíduos

Quando a condição essencial de independência dos dados é violada, ou seja, a presença de autocorrelação significativa é verificada não é possível implementar sem qualquer restrição os gráficos de controle abordados anteriormente.

Uma vez verificada a existência de autocorrelação significativa no processo, a implementação dos gráficos estatísticos prossegue com a determinação dos resíduos e/ou dos erros de previsão. Para isso é necessário determinar o modelo matemático que melhor se ajusta aos dados do processo.

Com base nestes resíduos será possível aplicar os diferentes tipos de gráficos de controle desde que nas fórmulas apresentadas se substituam os valores da característica da qualidade X pelos respectivos resíduos ou erros de previsão.

2.1.18 Capacidade do Processo

A análise da capacidade do processo consiste em comparar a distribuição de uma determinada característica do produto/serviço com as especificações estabelecidas (MONTGOMERY, 2004). Essa capacidade do processo reflete a variabilidade mínima devido a causas comuns, obtida após a eliminação das causas especiais.

Esta análise é uma parte vital de um programa de melhoria de qualidade e permite: prever se o processo é capaz de produzir de acordo com as especificações; ajudar a selecionar ou a modificar um processo ou uma máquina; selecionar fornecedores e reduzir a variação do processo.

Uma maneira de se analisar um processo estável é calcular a capacidade do processo, e para a medição dessa capacidade utiliza-se índices específicos. Os índices de capacidade de processo podem efetivamente resumir informações do processo de forma sucinta. Esses índices são parâmetros adimensionais dos quais determinam se um processo possui variabilidade suficientemente baixa para cumprir as especificações pré-estabelecidas (KANE, 1986).

Para utilizar os índices de capacidade é necessário que o processo esteja sob controle estatístico, que a variável de interesse tenha distribuição próxima da normal e que, no caso de especificações bilaterais, a média do processo esteja centrada entre os limites de especificação superior e inferior. A seguir serão descritos os índices C_p e C_{pk} .

2.1.19 Índice de capacidade potencial (C_p)

O índice de primeira geração C_p é definido para casos em que a especificação é bilateral e é calculado por

$$C_p = \frac{LSE - LIE}{6\hat{\sigma}} \quad (55)$$

onde: LSE : limite superior de especificação;

LIE : limite inferior de especificação;

$\hat{\sigma}$: desvio padrão estimado do processo.

Considera-se, em geral, que $C_p = 1,33$ é um valor aceitável para processos ou máquinas existentes quando a especificação é bilateral. No entanto, um bom desempenho do processo implica que a variabilidade seja bastante reduzida, ou seja, que a capacidade do processo seja a mais alta possível. A interpretação do índice baseia-se na seguinte análise:

- $C_p < 1$: a capacidade do processo é inadequada à tolerância exigida. Deve-se tentar outro processo mais adequado às especificações, tentar diminuir a variabilidade ou alterar as especificações do produto;

- $1 \leq C_p \leq 1,33$: a capacidade do processo está em torno da diferença entre as especificações. Deve-se tentar as mesmas alternativas do item anterior;

- $C_p > 1,33$: a capacidade do processo é adequada à tolerância exigida.

A razão C_p , não considera se o processo está centrado no alvo (valor nominal) ou não, ele mede simplesmente a dispersão seis-sigma no processo. Logo se deve utilizar somente com esta condição, de que o processo está centrado entre os limites de especificação.

2.1.20 Índice de performance (C_{pk})

O índice de performance permite avaliar se o processo está sendo capaz de atingir o valor nominal de especificação. O C_{pk} pode ser interpretado como uma medida da capacidade real, pois considera não só a dispersão como também a localização da média e é calculado por:

$$C_{pk} = \min \left[\frac{LSE - \bar{X}}{3\hat{\sigma}}, \frac{\bar{X} - LIE}{3\hat{\sigma}} \right] \quad (56)$$

Onde: *LSE*: limite superior de especificação;

LIE: limite inferior de especificação;

$\hat{\sigma}$: desvio padrão estimado do processo;

\bar{X} : média do processo.

Quando a média do processo coincide com o valor nominal da especificação, tem-se $C_p = C_{pk}$. No caso em que existe apenas um limite de especificação, pode-se calcular:

- Existência apenas do limite inferior:

$$C_{pi} = \frac{\bar{X} - LLE}{3\hat{\sigma}} \quad (57)$$

- Existência apenas do limite superior:

$$C_{ps} = \frac{LSE - \bar{X}}{3\hat{\sigma}} \quad (58)$$

2.1.21 ARL em controle e o ARL fora de controle

Duas medidas de desempenho que são utilizadas para avaliar a qualidade dos gráficos de controle são *ARL em controle* (*Average Run Length in control*) e o *ARL fora de controle* (*Average Run Length out of control*). O *ARL em controle* representa o número de amostras observadas do processo até que um alarme falso seja detectado, já o *ARL fora de controle* representa o número de amostras observadas do processo até que um alarme verdadeiro seja detectado. Essas medidas são definidas, respectivamente por:

$$ARL_{Em\ Controle} = ARL_{EC} = \frac{1}{\alpha} \quad (59)$$

onde α representa a probabilidade de um ponto exceder os limites de controle. Para uma situação em que foi utilizado 3σ , a probabilidade de um ponto qualquer se situar entre os limites de controle é de 99,73% e conseqüentemente existe um risco α igual a 0,27% de um ponto pertencente à distribuição normal estar fora dos limites de controle. Assim, o valor correspondente será $\alpha = 0,0027$ e, no processo sob controle, tem-se:

$$ARL_{EC} = \frac{1}{\alpha} = \frac{1}{0,0027} = 370$$

O valor 370 indica que, em média, a cada 370 pontos existirá um falso alarme, portanto $ARL_{EC} = 370$ é o número esperado de pontos registrados até que ocorra um alarme falso no processo.

Considerando agora que β é a probabilidade de se cometer o erro tipo II, ou seja, de se assumir que o processo está sob controle quando na realidade não está, tem-se:

$$ARL_{Fora\ de\ Controle} = ARL_{FC} = \frac{1}{1-\beta} \quad (60)$$

Nesse caso, o valor de ARL_{FC} é o número médio de observações que devem ser retiradas até a detecção de um verdadeiro deslocamento no processo, uma vez que ele já tenha ocorrido (ARAÚJO, 2000).

O ARL tem sido a medida mais utilizada para avaliar o desempenho de uma carta de controle e a análise conjunta do ARL_{EC} e ARL_{FC} indica a eficácia dessas cartas. Quando o processo se encontra sob controle estatístico é desejável que o seu valor seja o maior possível - minimizando o número de falsos alarmes - e o menor possível quando se encontra fora de controle - maior rapidez em detectar uma alteração do parâmetro do processo (PEREIRA; REQUEIJO, 2008).

A avaliação das curvas características de operações foge ao escopo pelo fato de que os dados a serem analisados não são provenientes de um processo produtivo.

2.1.22 Função de autocorrelação

i) Função de autocorrelação (FAC): Chama-se autocorrelação de defasamento k ("lag k ") à correlação entre quaisquer duas observações defasadas de k instantes. Esta autocorrelação é definida através do coeficiente de correlação ρ_k , ou seja, a autocorrelação entre x_t e x_{t-k} é dado por:

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{V(X_t)} \quad , k = 0, 1, 2, \dots \quad (61)$$

onde: $Cov(X_t, X_{t-k})$ é a covariância de observações defasadas em k instantes;

$V(X_t)$ é a variância de X .

A sequência de pares (k, ρ_k) é denominada função de autocorrelação. Valores negativos de k não são considerados explicitamente, tendo em vista que $\rho_k = \rho_{-k}$.

O coeficiente de autocorrelação ρ_k envolve parâmetros geralmente desconhecidos. Na prática, é necessário trabalhar com o coeficiente de autocorrelação amostral r_k para estimar ρ_k , expresso por:

$$r_k = \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (62)$$

onde n é o número de observações da série x_t .

A FAC amostral é constituída pelos pares (k, r_k) , $k = 1, 2, \dots$

O conjunto formado pelos coeficientes de correlação ρ_k , para $k = 0, 1, 2, \dots$, designa-se por FAC.

Para verificar se os dados referentes à variável X são autocorrelacionados, deve-se construir o gráfico da FAC ($\hat{\rho}$ em função de k) conhecido como correlograma. O critério de decisão consiste em verificar se todos os valores de r_k pertencem a um determinado intervalo de confiança, o que corresponde à ausência previsível de autocorrelação dos dados.

Este intervalo de confiança é determinado com base no valor esperado e na variância de r_k . Como r_k tem uma distribuição aproximadamente Normal ($r_k \sim \mathcal{N}(0, Var(r_k))$), o intervalo de confiança, para um nível de significância α , será dado pela equação:

$$-z_{\alpha/2}\sqrt{\hat{v}} \leq r_k \leq z_{\alpha/2}\sqrt{\hat{v}}, \quad \text{onde } \hat{v} = Var(r_k) \quad (63)$$

ii) Função de Autocorrelação Parcial

As séries temporais podem ser descritas através de modelos autoregressivos (AR), de médias móveis (MA) ou de uma mistura dos dois (ARMA), desde que o processo seja estacionário (processo com média e variância constantes ao longo do tempo e o valor da covariância dependendo apenas da defasagem). A modelação de um processo não estacionário pode ser realizada com recurso aos modelos ARIMA.

Segundo Pereira e Requeijo (2008), se o processo for descrito por um modelo AR, a FAC começa a decrescer a partir de determinada ordem de defasamentos, embora nunca atinja o valor zero. Já em modelos MA, a função de autocorrelação anula-se a partir de uma certa ordem de defasamento. Assim, pode-se perceber que o comportamento da FAC dificulta a escolha do melhor modelo para descrever o

processo (baseado na série temporal), sendo necessário o uso de outro instrumento, a FACP.

A autocorrelação parcial com defasagem k é definida como a correlação entre X_t e X_{t+k} com os efeitos das observações $(X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$ removidos. Em notação, o coeficiente de autocorrelação parcial de ordem k é usualmente designado por ϕ_{kk} .

Os coeficientes ϕ_{kk} não são conhecidos e portanto devem ser estimados. O conjunto formado pelos coeficientes de autocorrelação parcial estimados $\hat{\phi}_{kk}$ constitui a função de autocorrelação parcial.

Para verificar se a autocorrelação parcial para o instante k tem um efeito significativo, é necessário testar se os valores de ϕ_{kk} são significativamente diferentes de zero. Para tanto, calcula-se o intervalo de confiança para um nível de significância α . Este intervalo de confiança é função do valor esperado e da variância de $\hat{\phi}_{kk}$. Se o processo for autorregressivo de ordem p , os coeficientes de autocorrelação seguem aproximadamente uma distribuição Normal com média zero e variância $Var(\hat{\phi}_{kk})$ (VASCONSELOS; ALVES, 2000).

2.1.23 Modelagem de séries temporais de acordo com Box e Jenkins

A publicação por Box, Jenkins e Reinsel de *Time series analysis: forecasting and control* (1994) conduziu a uma nova geração de ferramentas de previsão. Popularmente conhecida como metodologia de Box-Jenkins, mas tecnicamente como metodologia ARIMA, esses modelos resultam da combinação de três componentes também denominados “filtros”: o componente Autorregressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA).

Uma série de tempo pode conter os três componentes ou apenas um subconjunto deles, resultando daí várias alternativas de modelos passíveis de análise pela metodologia de Box-Jenkins, conforme será visto a seguir.

2.1.24 Processo estacionário

Considera-se que um processo é estacionário quando ao realizar um mesmo deslocamento no tempo de todas as variáveis de qualquer distribuição conjunta finita, resulta que esta distribuição não varia, isto é, a distribuição de N observações a partir de um determinado instante t ($X_t, X_{t+1}, \dots, X_{t+N-1}$) tem as mesmas propriedades estatísticas que a distribuição de N observações defasadas de k instantes ($X_{t+k}, X_{t+k+1}, \dots, X_{t+k+N-1}$). Um processo é não estacionário quando não se verifica a condição enunciada anteriormente. A defasagem no tempo k é responsável por uma alteração no valor de um ou dos dois parâmetros do processo (média e variância).

Dessa forma, segue que quando um processo é estacionário, o ajuste da série temporal é realizado por um modelo ARMA (p,q). Se o processo não é estacionário, é necessário aplicar o operador de diferenças no sentido de transformar a variável original X em uma variável estacionária Y definida, no instante t , por $Y_t = \nabla^d X_t$.

A partir do momento em que uma série temporal é estacionária, pode-se modelá-la com uma variedade de formas e algumas delas serão abordadas a seguir.

2.1.25 Autocorrelação

A autocorrelação pode ser definida como a correlação entre integrantes de séries de observações ordenadas no tempo (como as séries temporais) ou no espaço (como nos dados de corte transversal).

Segundo Gujarati e Porter (2011), o modelo clássico de regressão linear pressupõe que o termo do erro relacionado a qualquer uma das observações não é influenciado pelo termo do erro de qualquer outra observação, conseqüentemente a autocorrelação não existe em termos de erro a_t . Simbolicamente

$$\text{cov}(a_t, a_{t+1} | x_t, x_{t+1}) = E(a_t, a_{t+1}) = 0 \quad (64)$$

Contudo, se for verificada essa dependência, teremos autocorrelação, ou seja

$$E(a_t, a_{t+1}) \neq 0 \quad (65)$$

Muitas vezes a correlação serial ocorre em estudos de séries temporais quando os erros associados às observações em um dado período de tempo se mantêm por transferência nos períodos futuros. Isso pode acontecer ocasionalmente com dados em corte transversal quando as unidades de observação têm uma ordem natural (PINDYCK; RUBINFELD, 2004).

Assim, em situações onde a condição essencial de independência dos dados é violada, ou seja, a presença de autocorrelação significativa é verificada, não é possível implementar sem qualquer restrição os gráficos de controle abordados anteriormente.

Dentre as formas adequadas para verificar a existência de autocorrelação significativa (ou verificar se os dados não são independentes), a Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) são alguns dos instrumentos utilizados para esta identificação.

2.1.26 Modelos Univariados

i) Modelo autorregressivo de ordem AR (p)

De acordo com esse modelo, y_t é descrito apenas por seus valores passados e pelo ruído branco ε_t . A forma mais simples de um modelo AR ocorre quando y_t depende somente de y_{t-1} e de ε_t sendo denominado de processo autorregressivo estocástico de primeira ordem e descrito por:

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (66)$$

onde ϕ é um parâmetro desconhecido, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ e $E(\varepsilon_t \varepsilon_s) = 0$ para $t \neq s$.

Por se tratar de um modelo estacionário, a variância de $y_t(\gamma_0)$ deve ser constante e as autocovariâncias (γ_k) devem ser independentes de t dadas por:

Variância do AR(1)

$$\gamma_0 = E(y_t^2) = E(\phi y_{t-1} + \varepsilon_t)^2 = \phi^2 V(y_{t-1}) + V(\varepsilon_t) \quad (67a)$$

$$\gamma_0 = \phi^2 \gamma_0 + \sigma_\varepsilon^2 \quad (67b)$$

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1-\phi^2} \quad (67c)$$

Para que a variância de y_t seja não negativa e finita, é necessário que $|\phi| < 1$. Essa restrição imposta sobre o parâmetro ϕ_1 é chamada condição de estacionariedade.

Autocovariância de ordem k

$$\gamma_1 = E(y_t y_{t-2}) \quad (68a)$$

$$\gamma_1 = \phi^k \gamma_0 \quad (68b)$$

Observa-se, portanto, que as autocovariâncias não dependem de t e sim de k . Como $|\phi| < 1$, pela condição de estacionariedade, quanto maior for o valor de k , ou seja, quanto maior for a distância entre as observações, menor a autocovariância.

O processo autorregressivo genérico, representado por AR(p) pressupõe que seja o resultado da soma ponderada de seus p valores passados, além do ruído branco ε_t :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (69)$$

ii) Modelo de médias móveis MA (q)

Por este processo, a série y_t resulta da combinação linear de termos de erro de ruído branco ocorridos no período corrente e nos períodos passados. O modelo é dado por:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad (70)$$

onde θ é um parâmetro.

A média de y_t é zero, já que ε_t é um ruído branco. Sua variância é:

$$\gamma_0 = E(y_t^2) = E[(\varepsilon_t - \theta \varepsilon_{t-1})^2] = E[\varepsilon_t^2] + \theta^2 E[\varepsilon_{t-1}^2] - 2\theta E[\varepsilon_t \varepsilon_{t-1}]$$

$$\gamma_0 = \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2$$

As autocovariâncias do MA(1) são definidas da seguinte forma:

$$\gamma_1 = E[y_t y_{t-1}] = E[(\varepsilon_t - \theta \varepsilon_{t-1})(\varepsilon_{t-1} - \theta \varepsilon_{t-2})]$$

$$\gamma_1 = -\theta \sigma_\varepsilon^2$$

$$\gamma_2 = E[y_t y_{t-2}] = E[(\varepsilon_t - \theta \varepsilon_{t-1})(\varepsilon_{t-2} - \theta \varepsilon_{t-3})]$$

$$\gamma_2 = 0$$

Percebe-se então que as autocovariâncias de ordem maior ou igual a 2 são nulas:

$$\gamma_k = 0, \quad k \geq 2$$

O modelo de médias móveis genérico envolve q valores defasados de ε e é indicado por MA(q) cuja equação é:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

iii) Modelo autorregressivo de médias móveis – ARMA(p, q)

Conforme o próprio nome indica, esse modelo é uma combinação dos dois anteriores: y_t apresenta um termos autorregressivos e termos de médias móveis q em que a especificação genérica apresenta a seguinte equação:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

iv) Modelos autorregressivos integrados de médias móveis – ARIMA(p, d, q)

Os modelos de séries temporais apresentados até então são baseados na suposição de que as séries temporais envolvidas são (fracamente) estacionárias, ou seja, apresentam de forma geral média e variância constantes e covariância invariante no tempo. Sabe-se, no entanto, que muitas séries temporais são não estacionárias, isto é, são integradas e necessitam ser diferenciadas d vezes para tornarem-se estacionárias.

Assim, se y_t é não estacionária mas $x_t = \nabla y_t = y_t - y_{t-1}$ é estacionária, então y_t é dita integrada de ordem 1. Se y_t precisar de duas diferenças para ser estacionarizada, ou seja, se $z_t = \nabla^2 y_t = \nabla(\nabla y_t) = \nabla(y_t - y_{t-1})$ é estacionária, então y_t é integrada de ordem 2.

Após diferenciar uma série temporal d vezes para torná-la estacionária e aplicar-lhe o modelo ARMA (p,q) , diz-se que a série temporal original é ARIMA (p,d,q) , ou seja, ela é uma série temporal autorregressiva integrada de médias móveis em que p denota os números de termos autorregressivos, d o número de vezes que a série deve ser diferenciada antes de tornar-se estacionária e que q o número de termos de média móvel. Esse modelo pode ser representado genericamente por:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

onde $w_t = \nabla^d y_t$.

v) Modelo Sazonal (SARIMA)

A sazonalidade refere-se às flutuações periódicas dentro do período analisado. As séries sazonais apresentam um importante tipo de correlação verificada entre s instantes de tempo, onde s é o número de observações contidas em um ano (por exemplo $s = 12$ para dados mensais e $s = 4$ para dados trimestrais).

Os modelos SARIMA são, na verdade, extensões dos modelos ARIMA com uma componente sazonal. Os modelos podem ser puramente sazonais, ou seja, modelos onde correlações existentes ocorrem entre instantes de tempo múltiplos de s ou os modelos podem ser apresentados além da correlação entre t e $t-s$, $t-2s$... a correlação entre tempos sucessivos recaindo no modelo sazonal multiplicativo geral denominado ARIMA $(p, d, q) \times (P, D, Q)_s$, cuja equação geral é dada por:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

onde $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o polinômio autorregressivo simples de grau p ;

$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$ é o polinômio autorregressivo múltiplo sazonal de grau P ;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o polinômio de média móvel simples de grau q ;

$\Theta(B^s) = 1 - \theta_1 B^s - \dots - \theta_Q B^s$ é o polinômio de média móvel múltiplo sazonal de grau Q ;

$\nabla^d \nabla_s^D y_t = w^t$ é o filtro não-linear aplicado à série original X_t que produz um processo estacionário w^t com $\nabla^d = (1 - B)^d$ e $\nabla_s^D = (1 - B^s)^D$;

ε_t é o ruído branco do modelo.

2.1.27 Etapas da metodologia Box-Jenkins

Segundo Gujarati e Porter (2011), o método consiste em quatro etapas:

Etapa 1. Identificação. Neste estágio descobre-se os valores apropriados de p, d e q . Nesta etapa o correlograma e o correlograma parcial auxiliam nessa tarefa.

Etapa 2. Estimação. Depois de identificados os valores apropriados de p e q , os próximo estágio é estimar os parâmetros dos termos autorregressivos e dos termos de média móvel incluídos no modelo. Às vezes, esse cálculo pode ser feito por mínimos quadrados simples, mas às vezes deve-se lançar mão de métodos de estimação não linear (nos parâmetros). Como essa tarefa é agora rotineiramente feita por pacotes estatísticos, a estimação acaba sendo sistemática.

Etapa 3. Verificação do diagnóstico. Após escolhido um modelo ARIMA ou SARIMA específico, e tendo seus parâmetros conhecidos, deve-se a seguir verificar se o modelo selecionado ajusta-se aos dados razoavelmente bem em detrimento de outros modelos. Esse é o motivo da modelagem ARIMA/Box-Jenkins ser mais arte do que ciência; uma habilidade considerável é requerida para escolher o modelo ARIMA correto. Um teste simples do modelo selecionado é verificar se os resíduos estimados com base nesse modelo são ruídos brancos, se forem, poderemos aceitar o ajuste específico; do contrário, deve-se recomeçar. Portanto, a metodologia Box-Jenkins é um processo iterativo.

Etapa 4. Previsão. Uma das razões da popularidade da modelagem ARIMA é seu sucesso na previsão a qual pode ser pontual ou intervalar e, no segundo caso, é preciso conhecer a distribuição do erro de previsão.

2.2 Vigilância Epidemiológica

A epidemia pode ser definida de uma maneira geral como sendo uma ocorrência incomum ou inesperada de uma doença. O termo tem origem grega e significa “sobre a população”. Nesse sentido, epidemiologia significa o estudo de alguma coisa que afeta a população observada (JEKEL, 1999).

A epidemiologia foi definida por Porta, Greenland e Last (2008) como “o estudo da ocorrência e distribuição de estados ou eventos relacionados com a saúde em populações específicas, incluindo o estudo dos determinantes que influenciam tais estados e a aplicação deste conhecimento para controlar os problemas de saúde”. A partir dessa definição pode-se perceber que os epidemiologistas estão preocupados não somente com a incapacidade, doença ou morte, mas também com a melhoria dos indicadores de saúde e com maneiras de promover saúde. Estes estudos epidemiológicos incluem vigilância, observação, testes de hipótese, pesquisas analíticas e experimentos.

No Brasil, o marco para institucionalização das ações de Vigilância Epidemiológica foi a Campanha de Erradicação da Varíola, 1966 a 1973. O modelo dessa campanha inspirou a Fundação de Serviços de Saúde Pública (FSESP) a organizar, em 1969, um sistema de notificação de determinadas doenças. O principal êxito relacionado a esse esforço foi o controle da poliomielite no Brasil, na década de 1980, que abriu perspectivas para a erradicação da doença no continente americano, finalmente alcançada em 1994 (BRASIL, 2005).

A lei nº 8.080/90 define vigilância epidemiológica no âmbito do Sistema Único de Saúde (SUS) como “um conjunto de ações que proporciona o conhecimento, a detecção ou prevenção de qualquer mudança nos fatores determinantes e condicionantes de saúde individual ou coletiva, com a finalidade de recomendar e adotar as medidas de prevenção e controle das doenças ou agravos”. Esse conceito, no entanto, passou por diversas mudanças ao longo do tempo. Arreza e Moraes (2010) apresentam, de maneira clara e dentro de um contexto histórico, a trajetória conceitual da vigilância em saúde.

A vigilância é uma característica essencial da prática epidemiológica e pode ser usada, segundo Bonita, Beaglehole e Kjellström (2010), para: identificar casos isolados ou agrupados; avaliar o impacto de eventos para a saúde pública e avaliar

tendências; medir fatores de risco para doenças; monitorar a efetividade e avaliar o impacto de medidas de prevenção e controle, estratégias de intervenção e mudanças nas políticas de saúde; e planejar e fornecer atenção aos doentes.

Assim, as funções da vigilância epidemiológica podem ser resumidas em: coleta de dados; processamento dos dados coletados; análise e interpretação dos dados processados; recomendação das medidas de controle apropriadas; promoção das ações de controle indicadas; avaliação da eficácia e efetividade das medidas adotadas e divulgação de informações pertinentes.

Segundo a Secretaria de Vigilância em Saúde (MINISTÉRIO DA SAÚDE, 2005), a vigilância epidemiológica tem como propósito fornecer orientação técnica permanente para os profissionais de saúde que têm a responsabilidade de decidir sobre a execução de ações de controle de doenças e agravos, tornando disponíveis, para esse fim, informações atualizadas sobre a ocorrência dessas doenças e agravos, bem como dos fatores que a condicionam, numa área geográfica ou população definida. Subsidiariamente, a vigilância epidemiológica constitui-se em importante instrumento para o planejamento, organização e operacionalização dos serviços de saúde, bem como a normatização das atividades técnicas correlatas.

A informação é instrumento essencial para a tomada de decisões em todos os ramos das atividades humanas. Para vigilância em saúde, ela se constitui no fator desencadeador do processo informação-decisão-ação.

No Brasil, o Ministério da Saúde (2002), destaca cinco sistemas de informação em razão da maior relevância para a vigilância: Sistema de Informações de Agravos de Notificação (SINAN); Sistema de Informações Sobre Mortalidade (SIM); Sistema de Informações Sobre Nascidos Vivos (SINASC); Sistema de Informações Hospitalares (SIH/SUS) e Sistema de Informações Ambulatoriais do SUS (SIA/SUS). O SINAN apresenta maior importância para a vigilância em saúde por ser o instrumento de informação das doenças de notificação compulsória de abrangência nacional.

A partir destes sistemas de informação existem diferentes formas de utilização da informação disponível para medir saúde e doença tais como: prevalência, incidência, mortalidade ou pode-se ainda usar simplesmente o número de casos existentes submetidos a alguma forma de monitoramento.

2.2.1 Prevalência

A prevalência (P) de uma doença é calculada como segue:

$$P = \frac{\text{Número de casos existentes}}{\text{número de pessoas na população}}$$

Nem sempre os dados sobre população em risco estão disponíveis. Por essa razão, em muitas situações, a população total da área estudada é utilizada como uma aproximação.

Segundo Bonita, Beaglehole e Kjellström (2010), a taxa de prevalência é frequentemente expressa como casos por 100 (%) ou por mil (‰) pessoas. Neste caso, “P” tem de ser multiplicado por 10ⁿ. Se o dado for coletado para um ponto específico de tempo, “P” é a “taxa de prevalência pontual”. Algumas vezes é mais conveniente utilizar a “taxa de prevalência no período”, calculada como o número total de pessoas que tiveram a doença em um determinado período de tempo, dividido pela população em risco de ter a doença no meio desse período. Sem levar em conta a idade das pessoas acometidas (ou em risco), os principais fatores que determinam a taxa de prevalência são:

- A severidade da doença (se muitas pessoas que desenvolvem a doença morrem, a prevalência diminui);
- A duração da doença (se uma doença é de curta duração, sua taxa prevalência é menor do que a de uma doença com longa duração);
- O número de novos casos (se muitas pessoas contraírem a doença, sua taxa de prevalência será maior do que se poucas pessoas a contraírem).

2.2.2 Incidência

A incidência refere-se à velocidade com que novos eventos ocorrem em uma determinada população. A incidência considera o período de tempo em que os indivíduos estão livres da doença, ou seja, em risco de desenvolvê-la. Para cada ano de observação, e até que a pessoa desenvolva a doença ou seja perdida do

acompanhamento, cada pessoa da população em estudo contribui com uma pessoa-ano ou dia, semana, mês no denominador. A taxa de incidência é calculada da seguinte forma:

$$I = \frac{\text{Número de casos novos em certos períodos}}{\text{Número de pessoas sob risco ao mesmo tempo}}$$

O numerador refere-se estritamente à primeira manifestação da doença. A unidade da densidade de incidência deverá incluir sempre uma dimensão de tempo (dia, mês, ano, etc.).

Para cada indivíduo na população, o tempo em risco é aquele durante o qual a pessoa permaneceu livre da doença. Para o cálculo da densidade de incidência, o denominador é constituído pela soma de todos os períodos livres de doença para todos os participantes do estudo.

2.2.3 Mortalidade

Os epidemiologistas utilizam com grande frequência as estatísticas de mortalidade para avaliar a carga de doença nas populações e, também, para avaliar mudanças na ocorrência de doenças ao longo do tempo. No entanto, este tipo de dado é fortemente afetado por várias fontes de erro podendo acarretar em estimativas errôneas devendo, por tanto, ser usado com cautela. O cálculo da taxa de mortalidade é dado por:

$$\text{Taxa de mortalidade geral} = \frac{\text{Número de óbitos no período}}{\text{População estimada do período}}$$

2.2.4 Uso dos Gráficos de controle em estudos Epidemiológicos

Rich e Terry, em 1946, propuseram pela primeira vez o uso dos gráficos de controle no estudo da incidência semanal de casos de poliomielite na Califórnia

(EUA), no período de 1929 a 1943, inclusive. Neste estudo foi usado os gráficos de controle para variáveis (SHEWHART, 1931).

Para o cálculo do limite superior foram utilizados dados dos anos 1931, 1932 e 1933 por apresentarem um comportamento padronizado (sem períodos epidêmicos), onde obteve-se média 4 e LSC igual a 10. Uma epidemia era caracterizada quando a incidência semanal se mantivesse acima do LSC por duas ou mais semanas sucessivas.

Em uma análise preliminar dos dados, os autores verificaram também a existência de sazonalidade ou seja, as epidemias se iniciavam na maioria das vezes entre as semanas 21 e 40 (aproximadamente entre final de maio a final de setembro).

O LSC não epidêmico da incidência semanal de poliomielite manteve-se constante para todo o período analisado (Figura 2), o que torna este diagrama de controle limitado ao descrever padrões esperados de agravos com comportamento sazonal. Como alternativa, os autores sugerem o uso de técnicas como médias móveis ou análise através de séries temporais que incorporem funções de seno e cosseno.

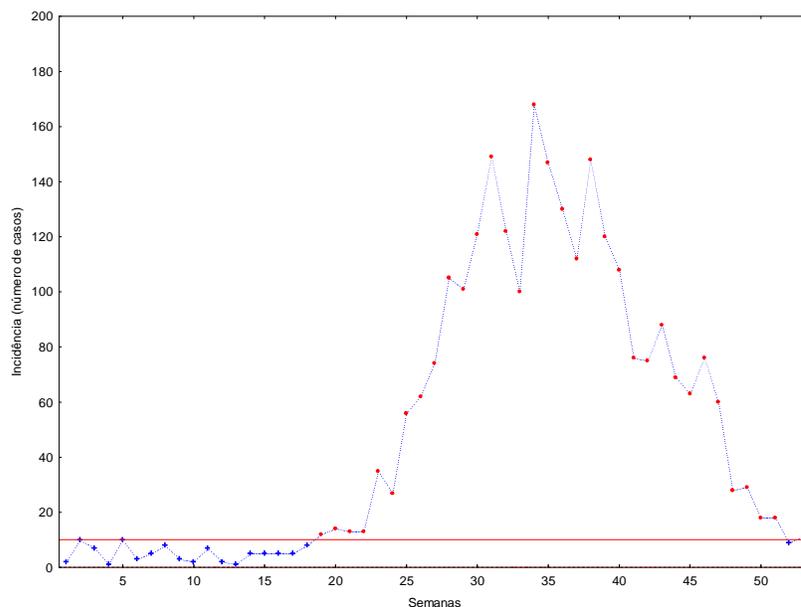


Figura 2 - Diagrama de controle para incidência de poliomielite, Estado da Califórnia, EUA, 1943.
Fonte: Rich e Terry, 1945.

Rich e Terry relatam ainda que outros refinamentos são necessários com intuito de eliminar o efeito de um longo tempo de tendência (secular), que pode ser ocasionado, por exemplo, pelo aumento da população. Schilling (1944) apresenta métodos de estimação de sazonalidade e tendência secular empregados antes da construção do gráfico de Shewhart, no controle do número de visitas realizadas por enfermeiros da Associação de Enfermagem de São Francisco (EUA).

Cullen et al. (1984) fizeram uso de diagramas de controle com o intuito de implantar um sistema de alerta precoce de surtos epidêmicos de malária em duas regiões no norte da Tailândia. Esse sistema deveria ser eficaz o suficiente para permitir a aplicação mais racional de controle de medidas de prevenção, ao mesmo tempo que fosse consideravelmente simples para ser aplicado no nível mais baixo dos setores de organização de controle da doença. O método a ser implementado também deveria ser confiável o bastante para indicar o número anormal de casos em áreas limitadas, e com nível de sensibilidade satisfatório para dar avisos oportunos de surtos iminentes.

Os dados de incidência de malária usados no trabalho foram coletados mensalmente nos períodos de 1973 a 1981. Quatro métodos diferentes de obtenção da média do processo de controle foram descritos: média geral, média móvel, média acumulada e média mensal, a qual foi eleita pelos autores para o desenvolvimento dos diagramas de controle apresentados no estudo. O cálculo do limite superior foi baseado nos anos iniciais os quais não apresentaram episódios de epidemia, isto é, sem observações que ultrapassassem 4 desvios padrão em relação média mensal. Tomou-se como base a distribuição normal, porém com um intervalo de 95% de confiança.

Braz (2005) comparou cinco diferentes gráficos de controle na detecção precoce de epidemias de malária em três diferentes localidades da Região da Amazônia Legal do Brasil no período de 1996 a 2003. Para a comparação entre os métodos empregados, foi considerado como mais adequado o método estatístico que apresentou maior taxa de alarmes verdadeiros de meses epidêmicos (TAV) disparados por cada método.

Esse indicador foi adotado visando à identificação do método mais sensível para detecção de verdadeiras epidemias, evitando a ocorrência de dois tipos de erros: I e II. O erro de tipo I equivale ao falso positivo, ou seja, aceitar a existência de uma epidemia quando ela realmente não existe. O erro de tipo II equivale ao falso

negativo, rejeitar a existência de uma epidemia quando ela, realmente, existe. Para tanto, também foi utilizado o intervalo de 95% de confiança.

As infecções hospitalares configuram importante problema de saúde pública mundial pois causam aumento na morbidade, na mortalidade e no tempo de internação dos pacientes; também acarretam mudança nos padrões de resistência microbiana e conseqüente elevação nos custos assistenciais. Assim, os diagramas de controle em vigilância hospitalar também são discutidos na literatura.

Sellick (1993) apresentou um estudo de monitoramento de ocorrências de patógenos resistentes em infecções hospitalares usando gráficos para atributos. O gráfico do tipo *c* foi usado para monitorar a presença da toxina *Clostridium difficile* em exames mensais de pacientes. O gráfico do tipo *u* foi usado para controlar infecções nosocomiais em pacientes internados na unidade de terapia intensiva e o gráfico tipo *p* foi usado para inspeção de taxas de infecções cirúrgicas. Em todos os casos foi usado dois desvios padrão acima da média para limite de alerta e três desvios para LSC.

Benneyan (1998) discute a aplicação dos gráficos de controle na área da saúde. O autor afirma que nos dados de saúde, assim como nos industriais, não deve haver evidências de comportamento não-aleatório dentro dos limites tais como tendências e ciclos. Sob a filosofia do controle de processo estatístico, um primeiro passo para a redução da taxa nosocomial é levar o processo a um estado de controle estatístico de modo que se opere apenas com a variabilidade natural. A partir de dados sobre infecção no trato urinário associadas ao cateter, a construção e peculiaridades dos gráficos para atributos (*c*, *p* e *np*) são descritas no trabalho.

Araújo (2000) abordou o problema do controle da infecção hospitalar causado pelo MRSA (*Staphylococcus aureus* resistente à metilina) sob a utilização dos gráficos *x* e *R* (medidas individuais e amplitude móvel) e CUSUM (soma cumulativa) em dados coletados junto ao Hospital Universitário de Brasília e Hospital de Base de Brasília (Brasil). Através da interpretação visual os CUSUMs foram mais eficientes para detectar sinais de alerta.

Arantes et al. (2003) monitoraram a tendência de ocorrência de infecções hospitalares em pacientes internados em uma Unidade de Terapia Intensiva Pediátrica de um hospital universitário durante os anos de 1998 a 2000. Por meio da aplicação do gráfico do tipo *u* (para taxas de ocorrência por período) os autores

puderam distinguir causas incomuns (surtos ou epidemias) entre as variações naturais nas taxas de ocorrência de infecções hospitalares.

Gomes, Mingoti e Oliveira (2011) avaliaram casos de infecções hospitalares que foram registrados na unidade de cuidados intensivos em um hospital universitário de Belo Horizonte (Brasil) em 2004 e 2005. Os resultados mostraram que o gráfico Shewhart tradicional é o método mais adequado para a monitorização de períodos com grandes desvios, enquanto os gráficos de EWMA e CUSUM foram melhores para controlar os períodos com desvios menores da taxa média de infecção.

A taxa de ocupação de leitos em hospitais é uma variável que também pode ser monitorada por meio de gráficos de controle. A baixa percentagem de ocupação torna o hospital deficitário economicamente porque continua com as mesmas despesas físicas diminuindo a receita. A ocupação muito elevada causa sobrecarga de trabalho na preparação de leitos, dificulta a higienização das enfermarias/leitos, prejudica o trabalho da farmácia e sobrecarrega as tarefas administrativas.

Burns et al. (2005) analisaram, ao longo de dois anos, a gestão de ocupação de leitos do Hospital da Universidade de Queensland situado na cidade de Brisbane (Austrália) através do método CUSUM. Este se mostrou ser um meio simples de revelar tendências importantes no fluxo de pacientes as quais não eram evidentes em observações convencionais em dados de séries temporais. Isso levou a otimização na administração dos leitos que resultou em uma diminuição de 9.500 ocupações ao longo de um ano.

Zanini (2006) investigou o potencial dos gráficos de controle no monitoramento de diferentes taxas de ocupação hospitalar. Para isso, utilizou alguns indicadores obtidos no Hospital Universitário de Santa Maria (Brasil), no período de 2000 a 2005: taxa de ocupação do hospital geral, taxa de ocupação do pronto atendimento, taxa de ocupação psiquiátrica, taxa de ocupação na Unidade de Tratamento Intensivo (UTI) de recém-nascidos além de trabalhar também com a taxa de mortalidade institucional por diálise. Nas séries que apresentaram correlação serial foram ajustados modelos ARIMA de Box e Jenkins. Comparando os gráficos utilizados, a autora verificou que os gráficos para medidas individuais x e R apesar de serem mais simples de serem implementados, são ineficientes para detectar pequenos desvios e insensíveis para emitir sinal de alerta preventivo. Já os EWMA

e, principalmente, os CUSUMs são mais eficientes em detectar pequenos desvios e capazes de manifestar sinais de alerta precoces mais rapidamente que os demais.

Um dos maiores objetivos em vigilância epidemiológica é a intervenção para reduzir a morbidade e mortalidade por doenças. Existe uma preocupação por parte dos epidemiologistas em desenvolver uma base racional para programas de prevenção. Assim, se é possível identificar os fatores etiológicos ou causais para a doença e reduzir ou eliminar a exposição a esses fatores, pode-se desenvolver uma base para programas de prevenção.

Dessa forma, procura-se melhorar a condição de saúde das populações pela eliminação e/ou atenuação dos riscos associados à rápida disseminação ou persistência com grande impacto sobre a morbimortalidade que as mesmas apresentam. Alguns trabalhos sobre o uso de gráficos de controle para variáveis de mortalidade em saúde pública estão disponíveis na literatura.

Rossi, Lampugnani e Marchi (1999) apresentam uma aproximação do processo CUSUM baseado na transformação da distribuição Poisson para a distribuição normal, o que pode ser uma solução eficiente para problemas previamente apontados nos dados. Para ilustrar essa aplicação foram utilizados dados de mortalidade por doenças respiratórias na região de Toscana (Itália) durante o período de 1980 a 1989.

Montoya-Restrepo e Correa-Morales (2009) aplicou a metodologia dos gráficos de controle (gráfico para proporções e logit das proporções) na vigilância de mortalidade perinatal na Colômbia. Com isso, pretendia vigiar a variação do processo reprodutivo ao longo do tempo e provar a efetividade das ações corretivas empregadas na melhoria da qualidade da atenção em saúde pública daquele país.

Williamson e Hudson (1999) usaram modelos Box-Jenkins associados aos gráficos de controle (Shewhart e EWMA) para detectar sinais de alerta na saúde pública nos Estados Unidos. Lima (2008) usou técnicas conjuntas de análise de séries temporais, análise de regressão e de controle estatístico de qualidade para o monitoramento de doenças de notificação compulsória na cidade de Belém-PA.

Shu, Jiang e Tsui (2011) compararam CUSUM ponderada (WCUSUM) e procedimentos CUSUM convencionais na presença de alterações no tamanho da população monótonas. Os resultados de simulação mostram que o método WCUSUM pode ser mais eficiente do que o convencional CUSUM métodos de detecção de aumentos na taxa de incidência, especialmente para pequenas

mudanças. Um exemplo baseado em dados de mortalidade do Novo México é utilizado para ilustrar a aplicação do método WCUSUM.

Muitos outros desfechos têm sido estudados com uso de técnicas de estatísticas de controle como por exemplo: aspectos epidemiológicos da meningite (FREITAS, MERCHÁN-HAMANN, 2004), mortalidade por cirurgia cardíaca (ROGERS et al, 2004; MOHAMMED, 2005; KUNADIAN, 2008; NOYEZ, 2009), sequenciamento genético (DAS, 2012).

É possível encontrar também *surveys* na literatura sobre o uso dos métodos de controle usados em vigilância epidemiológica: Costa (1994), Sanches (2000), Sonesson e Book (2003), Brookmeyer e Stroup (2004), Woodall (2006 e 2008) e Segó, Woodall e Reynolds Junior (2008) são alguns desses trabalhos.

3 METODOLOGIA

Neste capítulo serão apresentados os procedimentos metodológicos do estudo, assim descritos: métodos e técnicas utilizados, fontes de dados, caracterização e limitação da pesquisa e a estratégia analítica.

3.1 Fontes de dados

Para a análise do desempenho dos diferentes gráficos de controle estatístico foi usado, neste estudo, dados mensais de notificação de hepatite do tipo C, no Rio Grande do Sul (RS), proveniente de dados secundários obtidos no ambiente virtual do Sistema de Informação de Agravos de Notificação (SINAN). Os casos estão compreendidos no período de janeiro de 2007 a dezembro de 2011.

Utilizou-se como desfecho o número mensal de casos de hepatite C no estado do RS para construção dos gráficos referentes aos métodos estatísticos anteriormente abordados. Esse agravo foi eleito para ser analisado por se tratar de um importante problema mundial de saúde pública sendo a principal causa de doença hepática crônica, cirrose, hepatocarcinoma e transplante hepático.

No Brasil, aproximadamente 70.000 casos de hepatite crônica do tipo C foram confirmados entre os anos de 1999 e 2010. A taxa média de detecção de casos confirmados no Brasil, em 2011, foi de 5,0 casos por 100 mil habitantes, para a região Sul foi de 8,5 e para o Rio Grande do Sul, 10,2 (BRASIL, 2012).

No tratamento dos dados e aplicação dos testes estatísticos foram utilizadas as planilhas eletrônicas *Excel* e *Statistica* (versão 9.1).

3.2 Caracterização e limitação da pesquisa

Essa é uma pesquisa transversal, de caráter exploratório e com o objetivo de apresentar uma aplicação de gráficos de controle estatístico em dados epidemiológicos.

A utilização do SINAN como base de dados local é uma fonte valiosa de informações epidemiológicas, apesar de ainda existir falhas quanto ao preenchimento completo de todos os campos da ficha de notificação/investigação não configurando a realidade dos casos.

A gradual demora na notificação dos casos por parte dos sistemas de vigilância, bem como a falta de publicação de notificações diárias ou semanais também se torna um aspecto limitante do estudo, pois essa falta de alimentação nos bancos de dados específicos dificulta a implementação de métodos de controle e consequentemente, a detecção de surtos ou epidemias em tempo real.

3.3 Estratégia Analítica

Após a coleta, a elaboração e revisão do banco de dados, uma análise exploratória foi realizada sobre os dados de hepatite C, registrados no período de 2007 a 2011 para identificar as características da distribuição dos casos. Nesta análise, verificou-se a distribuição mensal e anual da doença no Rio Grande do Sul. Analisou-se também o ajuste da distribuição Normal esperada, dos casos da doença, por meio do teste Shapiro-Wilks.

Se a apresentou correlação serial, foram ajustados modelos de séries temporais usando a metodologia de Box-Jenkins para a posterior construção dos gráficos de controle nos resíduos da série.

A função de autocorrelação e autocorrelação parcial auxiliaram na identificação dos modelos e no número de parâmetros. Para eleger o melhor dentre os possíveis modelos identificados, foi utilizado o critério do menor valor de AIC e BIC. E com base nesse modelo procedeu-se a aplicação de três métodos estatísticos de controle: gráfico para medidas individuais, Gráfico de Controle das Somas Acumuladas (CUSUM) e Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA).

4 RESULTADOS

4.1 Análise Descritiva

Uma análise descritiva dos dados foi realizada, inicialmente, e na Tabela 1 constam os casos autóctones de hepatite C por mês e ano de ocorrência, acompanhados da média, desvio padrão e coeficiente de variação por período.

Pode-se verificar que apesar de apresentar um desvio-padrão elevado, as maiores médias de incidência da doença e os maiores coeficientes de variação ocorrem nos meses de Maio e Junho enquanto os menores valores para essas medidas ocorrem em Dezembro, Janeiro e Fevereiro, caracterizando a hepatite C como uma típica doença de inverno.

Tabela 1 - Casos autóctones de hepatite C, segundo o ano e o mês do Estado do Rio Grande do Sul, Brasil, 2007 a 2011.

Mês	Ano					Total	Média	Desvio-padrão	Coef.de Variação
	2007	2008	2009	2010	2011				
Jan	156	188	161	211	200	916	183,2	24,0	0,06
Fev	155	132	163	172	172	794	158,8	16,5	0,05
Mar	168	178	238	256	191	1031	206,2	38,7	0,08
Abr	146	188	220	201	175	930	186,0	27,9	0,07
Mai	192	195	261	351	221	1220	244,0	65,9	0,12
Jun	189	275	321	233	161	1179	235,8	64,6	0,12
Jul	166	185	237	214	202	1004	200,8	26,8	0,06
Ago	167	206	180	218	240	1011	202,2	29,3	0,06
Set	135	174	215	217	204	945	189,0	34,7	0,08
Out	171	159	232	204	231	997	199,4	33,9	0,08
Nov	123	174	191	191	242	921	184,2	42,7	0,10
Dez	107	174	204	218	182	885	177,0	42,8	0,11
Total	1875	2228	2623	2686	2421	11833	2366,6	24,0	-

Fonte: Ministério da Saúde, Secretaria de Vigilância em Saúde, Sistema de Informação de Agravos de Notificação (SINAN).

Para que os gráficos de controle estatístico tradicionais possam ser utilizados e, corretamente interpretados, é necessário que os dados analisados atendam a certos pressupostos como a independência e a normalidade da distribuição referente à estatística utilizada.

Para verificar se o ajuste da distribuição empírica dos casos de hepatite C se ajusta à distribuição teórica Normal, aplicou-se o teste de Shapiro-Wilks, no qual se constatou que a distribuição dos casos de hepatite não se ajusta a uma distribuição Normal, $p < 0,01$.

Sendo assim, partiu-se para a busca de uma abordagem apropriada para trabalhar com dados autocorrelacionados e não Normais.

4.2 Ajuste do modelo Box e Jenkins

Na Figura 3 representa-se o número de casos de hepatite C no Rio Grande do Sul, e com uma análise visual é possível verificar que existe tendência crescente e picos sazonais indicando que a série não é estacionária.

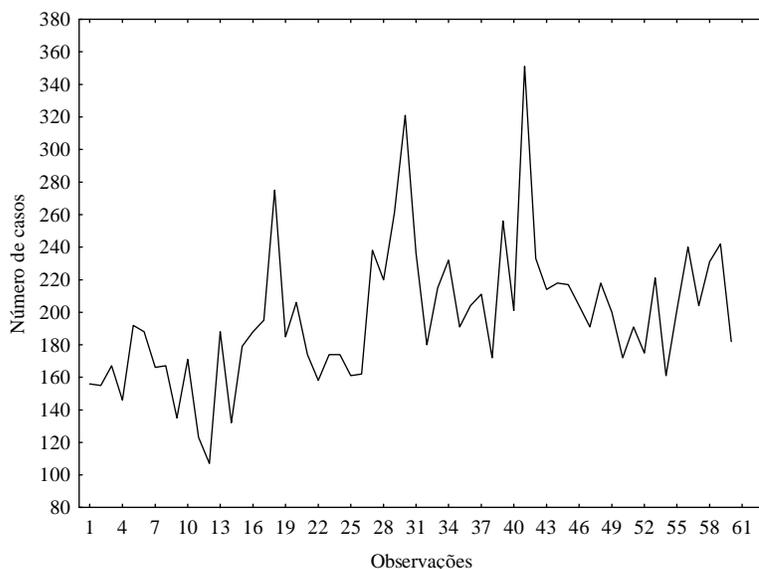


Figura 3 - Gráfico da série de casos mensais de Hepatite C do Rio Grande do Sul, nos anos de 2007 a 2011.

Com base nos correlogramas da função de autocorrelação e autocorrelação parcial também é possível verificar que não se trata de uma série temporal estacionária (Figuras 4 e 5). Assim se faz necessário uma diferenciação da série de forma a obter a estacionaridade pressuposta para a utilização da metodologia de Box e Jenkins (Figura 6).

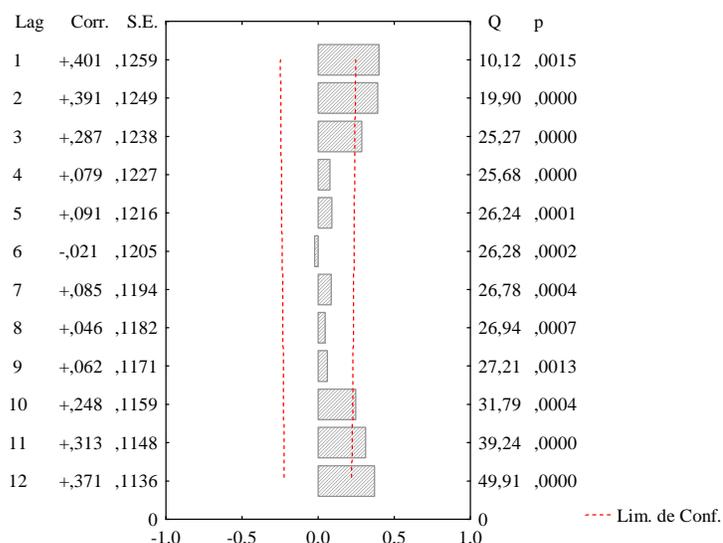


Figura 4 - Função de autocorrelação dos casos mensais de hepatite C, Rio Grande do Sul, 2007 a 2011.

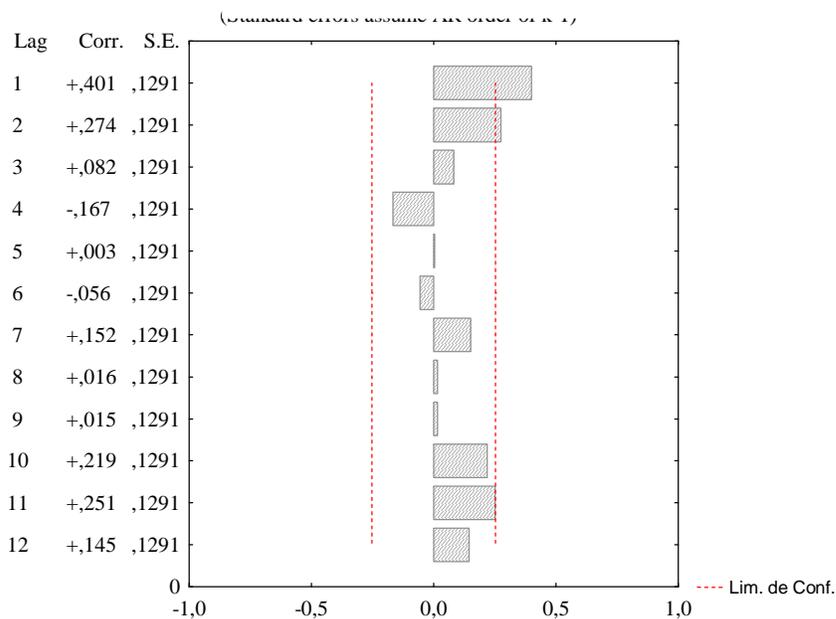


Figura 5 - Função de autocorrelação parcial dos casos mensais de hepatite C, Rio Grande do Sul, 2007 a 2011.

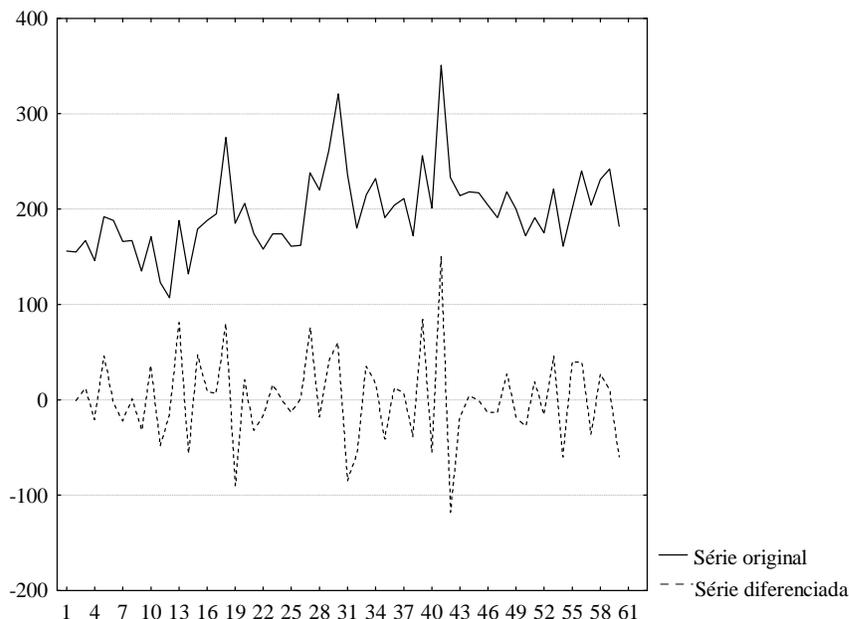


Figura 6 - Série histórica original e diferenciada dos casos de hepatite C. Rio Grande do sul, 2007 a 2011.

Pelos gráficos da função de autocorrelação e autocorrelação da série diferenciada, alguns modelos puderam ser ajustados e com o auxílio dos critérios dos menores valores de AIC e BIC um modelo $SARIMA(0,1,1) \times (1,0,0)_{12}$ foi escolhido, os demais modelos se encontram no Anexo D. Na Tabela 2 são apresentados os principais resultados do modelo escolhido.

Tabela 2 - Principais resultados do modelo $SARIMA(0,1,1) \times (1,0,0)_{12}$ ajustado para número de observações mensais de hepatite C no Rio Grande do Sul.

Ordem	Parâmetros	Erro padrão	Limite inferior (95%)	Limite superior (95%)	p-valor
q(1)	0,806	0,097	0,612	0,999	<0,001
Ps(1)	0,433	0,128	0,206	0,720	<0,001

O modelo pode ser escrito como $\Phi(B^s)\nabla^d y_t = \theta(B)\varepsilon_t$ e substituindo os parâmetros estimados tem-se:

$$\nabla^1 Y_t = \frac{0,806 \varepsilon_t}{0,433 Y_{t-12} + \varepsilon_t}$$

onde ε_t é o resíduo, com distribuição normal, independente e identicamente distribuído, portanto, um ruído branco, conforme mostram as Figuras 7 e 8.

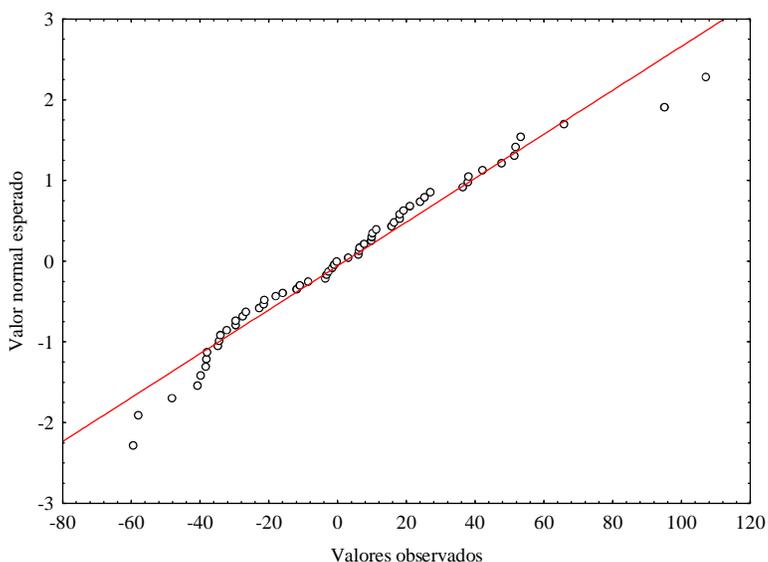


Figura 7 - Gráfico de normalidade dos resíduos para o modelo SARIMA (0,1,1)x(1,0,0)₁₂.

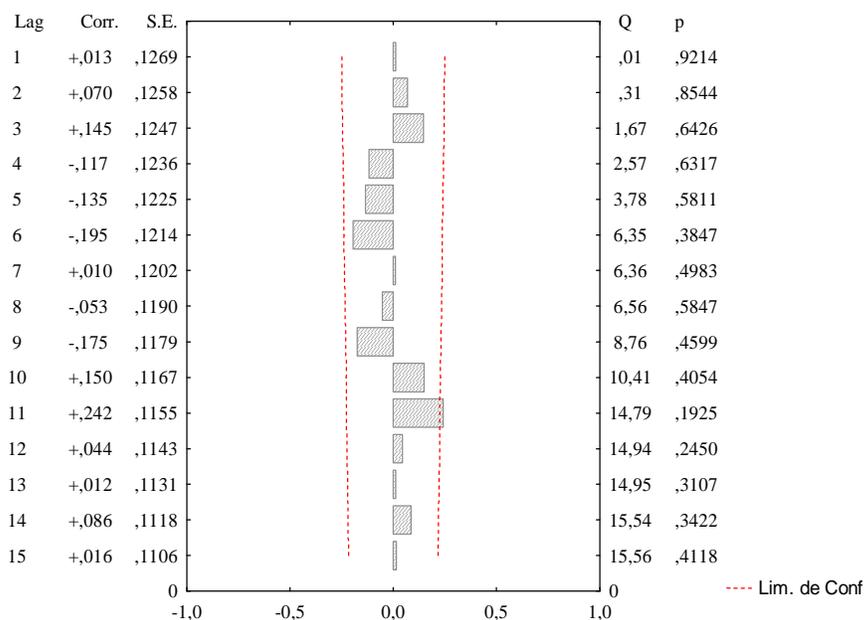


Figura 8 - Função de autocorrelação dos resíduos para o modelo SARIMA (0,1,1)x(1,0,0)₁₂.

A partir do ajuste de um modelo adequado e a constatação de que os resíduos se comportam como um ruído branco procedeu-se a construção dos gráficos de controle dos mesmos, conforme serão mostrados a seguir.

4.3 Aplicação dos Gráficos de Controle

4.3.1 Gráfico para medidas individuais

Os gráficos de controle foram implementados de acordo com os limites dos processos industriais, os quais utilizam na maioria dos casos a média e mais ou menos dois desvios-padrão (limites de alerta) ou 3 desvios-padrão (limites de controle). No processo de vigilância em saúde deseja-se identificar o quanto antes uma alteração no comportamento da doença, assim alguns trabalhos nessa área recomendam adotar dois desvios-padrão como limite de controle (ALVES, 2004).

Dessa forma, optou-se por plotar os gráficos com os limites de controle compreendidos em dois desvios-padrão devido a importância da detecção precoce de acometimentos da doença analisada.

Analisando-se o gráfico para medidas individuais (Figura 9) pode-se verificar que as observações 17 e 40 são identificadas como momentos de surto de hepatite na região Sul, considerando o limite de controle usando dois desvios-padrão, mesmo considerando as regras adicionais de interpretação gráfica (padrões especiais).

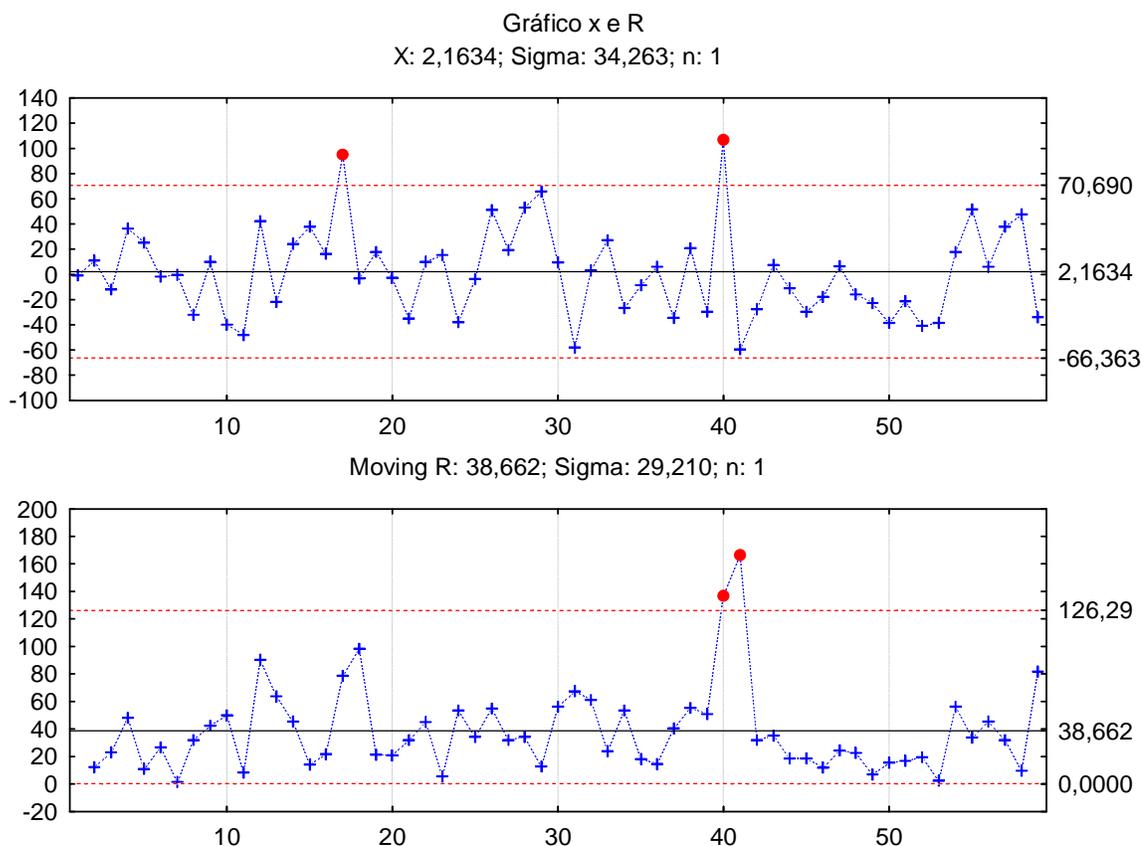


Figura 9 - Gráfico de controle aplicado aos resíduos do modelo para o número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.

4.3.2 Gráfico de CUSUM

Da mesma forma como foi feito para o gráfico das medidas individuais e médias móveis, os resíduos gerados pelo modelo SARIMA foram utilizados para a construção do Gráfico de Controle das Somas Acumuladas tabular (CUSUM tabular) conforme apresentado na Figura 10.

Neste caso, o planejamento do gráfico foi realizado seguindo as referências de Montgomery (2004) as quais definem o valor de sensibilidade (K) igual a 0,5 e intervalo de decisão (H) igual a 5. Estes valores remetem a um cenário de limites de controle semelhante ao do gráfico de controle de Shewhart com os limites 3σ habituais. Os cálculos para o CUSUM tabular se encontram no Anexo E.

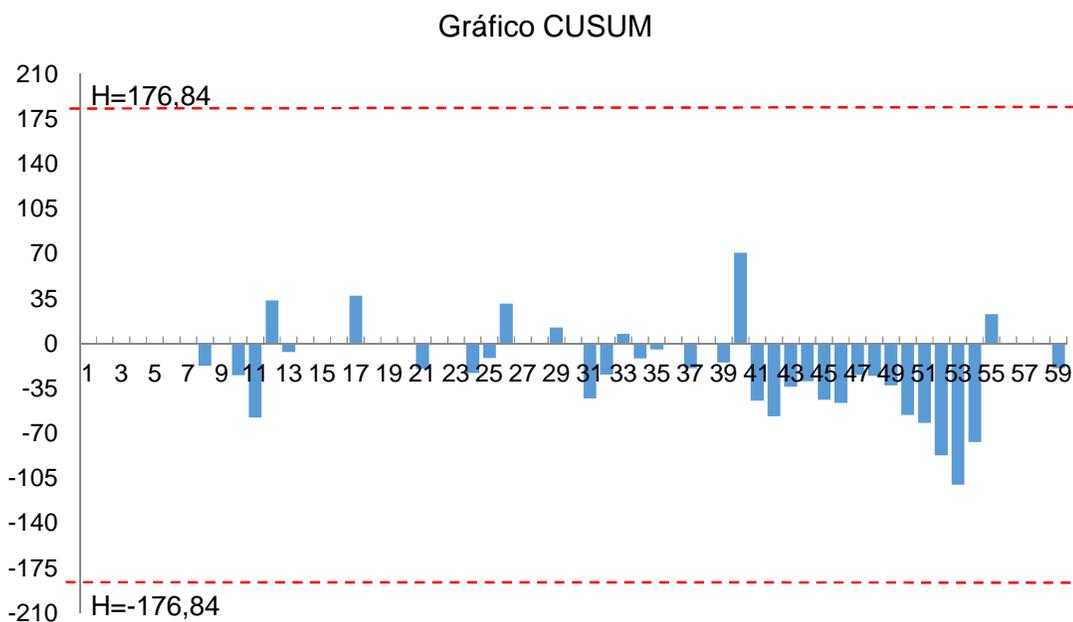


Figura 10 -Gráfico de controle com parâmetros $K=0,5$ e $H=5$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.

No CUSUM tabular cada barra representa o valor de C_i^+ e C_i^- versus o número da amostra. A ação tomada em seguida a um sinal de fora de controle em um esquema CUSUM é idêntica à de qualquer gráfico de controle. Os valores calculados para a plotagem deste gráfico se encontram no Anexo E.

Dessa forma, levando em consideração o intervalo de decisão determinado, o gráfico não permite detectar pontos fora do controle.

4.3.3 Gráfico de EWMA

A seguir apresenta-se o Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) para os resíduos do modelo. Adotou-se três valores para a constante ponderação (λ): 0,10, 0,40 e 0,90. Esses valores foram escolhidos com o propósito de avaliar este gráfico sob diferentes cenários acerca dos pesos associados às observações correntes e anteriores.

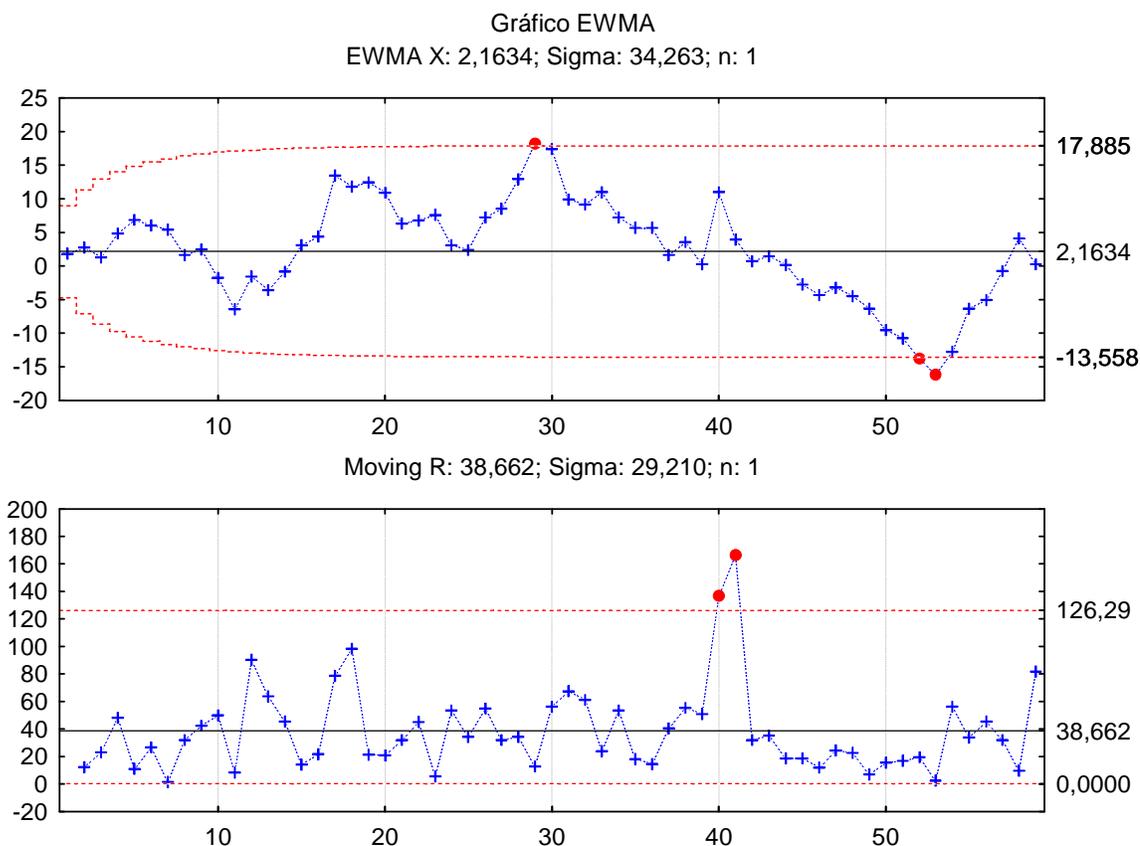


Figura 11 - Gráfico de controle com parâmetros $L=2$ e $\lambda=0,10$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.

Considerando $\lambda = 0,10$ (Figura 11) pode-se identificar apenas a observação 29 como fora do limite superior. Esta observação equivale, de fato, a um número elevado de casos de hepatite C na região Sul. A observação 30 se encontra muito próxima deste limite e também remete-se a um número elevado de casos da doença. Ainda neste gráfico percebe-se uma sucessão de pontos consecutivos e em declínio entre as observações 42 a 53, culminando com duas observações fora do limite inferior caracterizando a ação de uma causa especial, a qual, neste caso pode estar associada à uma campanha de combate à hepatite C lançada no final do mês de julho de 2010.

Na Figura 12 onde $\lambda = 0,40$, duas observações são plotadas fora do limite superior (17 e 29) enquanto os pontos 40 e 58 encontram-se nas proximidades deste limite. Da mesma forma que no caso anterior, ocorre uma sucessão de pontos em declínio finalizando com o ponto 53 fora do limite inferior.

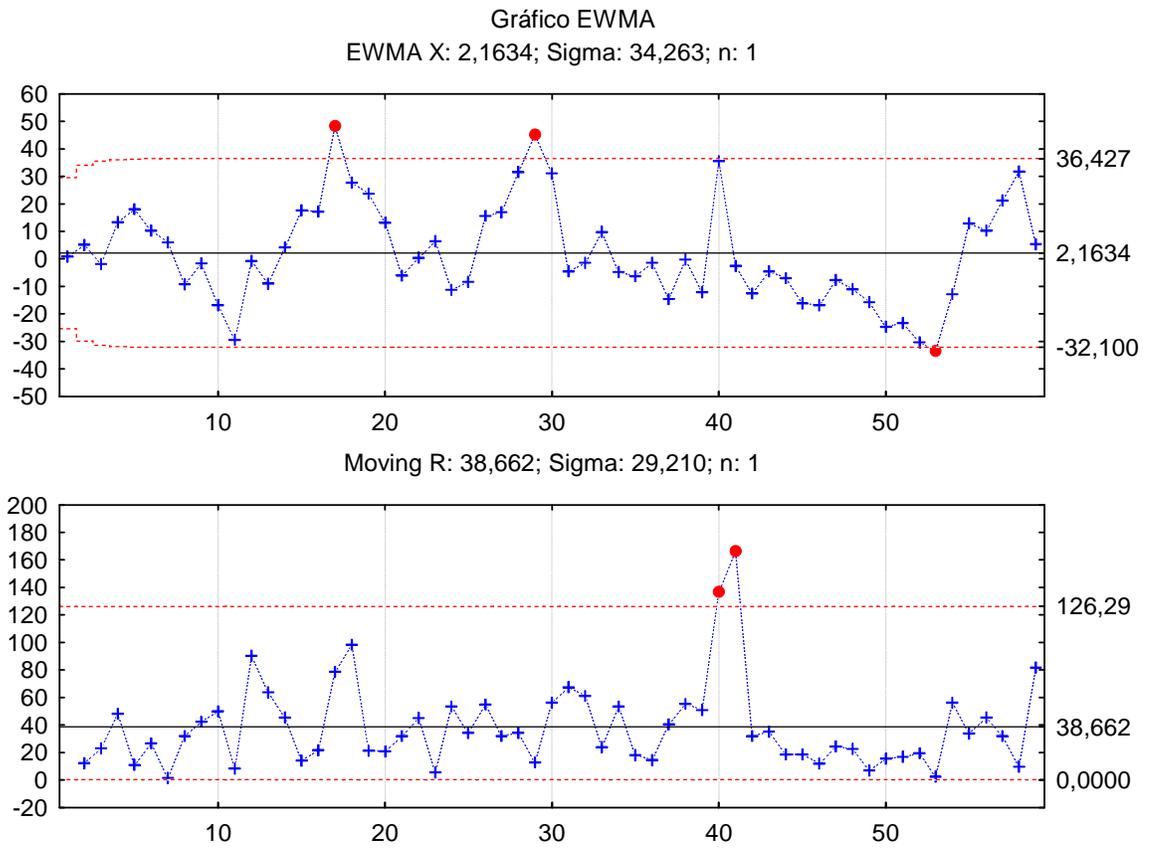


Figura 12 - Gráfico de controle com parâmetros $L=2$ e $\lambda=0,40$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.

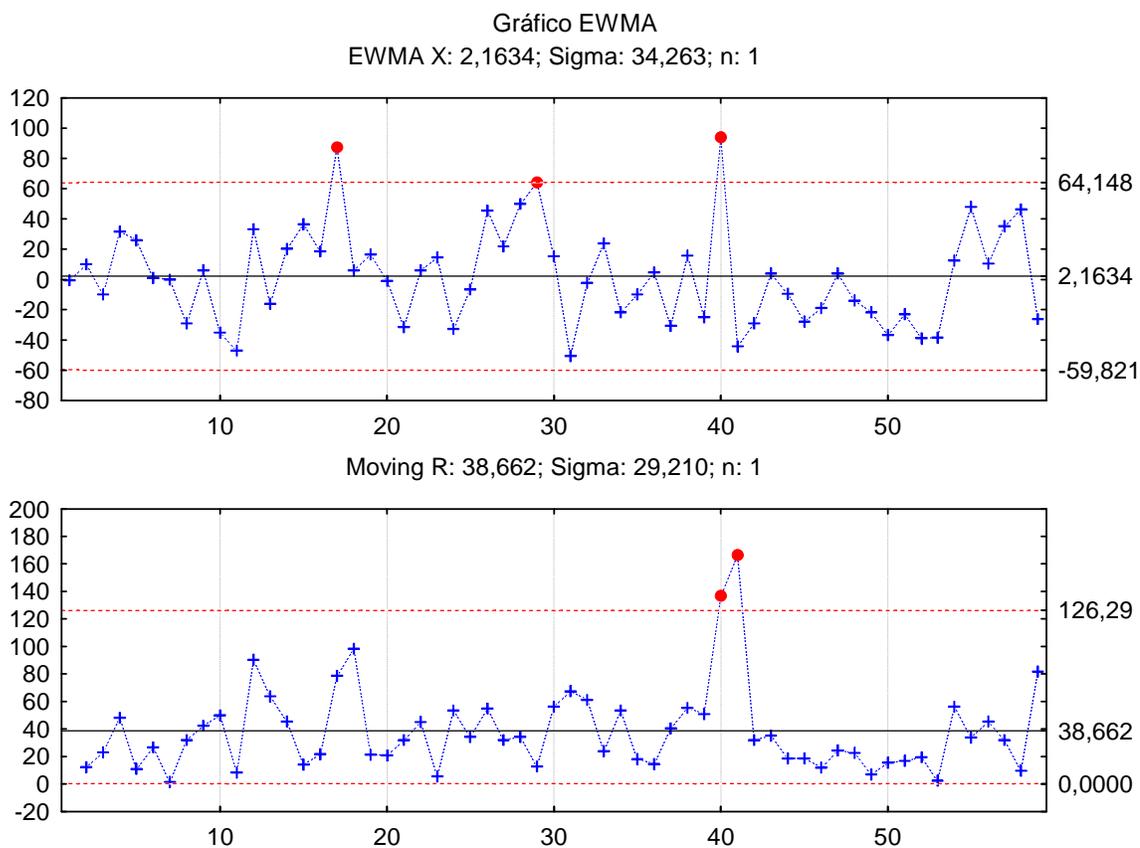


Figura 13 - Gráfico de controle com parâmetros $L=2$ e $\lambda=0,90$ aplicado aos resíduos do modelo do número de casos de hepatite C no Rio Grande do Sul no período de 2007 a 2011.

Analisando o último gráfico de EWMA, em que se considerou $\lambda = 0,90$, verifica-se mais pontos fora do limite superior (observações 17, 29 e 40).

Este cenário do gráfico de EWMA apresentou-se mais eficiente na detecção precoce dos pontos fora do limite.

5 CONCLUSÃO E CONSIDERAÇÕES FINAIS

Os casos de hepatite C no Rio Grande do Sul não apresentam características próximas da distribuição Normal, são autocorrelacionados e apresentam um período sazonal bem caracterizado.

As séries formadas por casos de doenças de notificação compulsória (ver anexo F), em geral, são semelhantes, em suas características, ao conjunto de dados apresentados neste trabalho. Dessa forma, tornam-se necessárias transformações na série para que possa ser adequadamente controlada pelos gráficos de controle.

A aplicação da metodologia de Box e Jenkins para séries temporais associadas à construção dos gráficos de controle podem constituir um importante suporte de orientação e apoio para o monitoramento de dados de vigilância epidemiológica.

O gráfico \bar{x} e R apresentou como vantagem a fácil elaboração, visualização e sensibilidade para detectar grandes mudanças na série. A implementação de limites de controle utilizando dois desvios padrão é possível sem que haja comprometimento da metodologia. No entanto, algumas observações consideradas importantes, devido ao grande número de casos de hepatite C, não foram detectadas por este método. Além disso, observações consideradas fora de controle são detectadas tardiamente.

A metodologia baseada no CUSUM, sensível na detecção de pequenos desvios, se mostrou ineficiente neste estudo devido ao fato de a série abordada apresentar uma maior variabilidade do que as variáveis normalmente apresentam. Também houve dificuldade em estabelecer os limites de controle para o uso equivalente a dois desvios padrão com base nos parâmetros tabelados.

O Gráfico de Médias Móveis Exponencialmente Ponderadas (EWMA) mostrou ser mais eficiente do que os gráficos de medidas individuais e amplitude móvel e das somas acumuladas para detecção de observações com grande número de casos de hepatite.

Este método apresenta uma fácil implementação de diferentes limites de controle e uma flexibilidade maior de seus parâmetros quando comparado ao gráfico

CUSUM. Com interpretação visual simples, foi mais sensível que os demais no que se refere às observações consideradas importantes. E dentre os cenários analisados para este gráfico, o caso em que $\lambda = 0,90$ mostrou ser superior aos demais em relação à detecção de surtos da doença analisada.

O uso de técnicas estatísticas para detecção de epidemias de doenças de notificação compulsória auxiliam os profissionais responsáveis pela condução das análises estatísticas rotineiras de dados de vigilância na identificação de padrões fora do controle das doenças.

Neste estudo foi possível verificar, com base na revisão de literatura, um número crescente de trabalhos com enfoque aplicado à prática da análise de dados de vigilância através de gráficos estatísticos de controle.

Assim, embora existam algumas dificuldades em se trabalhar com dados de vigilância epidemiológica (alimentação tardia dos bancos de dados, dados autocorrelacionados e/ou não normais, dificuldade em ajustar um modelo adequado), os gráficos de controle podem facilitar a análise de dados em saúde pública orientando gestores a tomar ações efetivas de prevenção e controle com base nos resultados observados.

É importante ressaltar que em vigilância de saúde o tempo é essencial para a detecção precoce de epidemias/surtos para que as medidas de controle sejam adotadas oportunamente, de modo que grande número de casos e óbitos possam ser prevenidos.

Para pesquisas futuras sugere-se:

- Aplicar os gráficos de controle aos dados de incidência da doença ou seja, o número de casos novos ocorridos em um certo período de tempo considerando o tamanho da população específica.
- Usar o procedimento combinado CUSUM-Shewhart para melhorar a sensibilidade do CUSUM para grandes mudanças.
- Monitorar a variabilidade do processo através do método CUSUM.
- Avaliar os resultados do monitoramento e controle de processos multivariados considerando dados subgrupados para uma mesma doença em diferentes regiões.

REFERÊNCIAS

ALVES, R.G. **Utilidade do Diagrama de Controle de Doenças na Previsão e Controle de Epidemias**. 116f. Dissertação (Mestrado em Saúde Coletiva) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

ALWAN, L. C.; ROBERTS H. V. Time-Series Modeling for Statistical Process Control. **Journal of Business & Economic Statistics**, v. 6, n. 1, p. 87-95, jan. 1988.

ARANTES, A. et al. Use of statistical process control charts in the epidemiological surveillance of nosocomial infections. **Revista de Saúde Pública**, v.37, n.6, p.768-774, 2003.

ARAÚJO, A. A. **Controle de qualidade em Epidemiologia**: aplicação de gráficos de controle para medidas individuais no controle de infecção hospitalar. 131f. Dissertação (Mestrado em Estatística e Métodos Quantitativos) – Universidade de Brasília, Brasília, 2000.

ARREAZA, A. L. V; MORAES, J. C. Vigilância da saúde: fundamentos, interfaces e tendências. **Ciência & Saúde Coletiva**, v. 15, n. 4, p. 2215-2228, 2010.

BENNEYAN, J. C. Statistical quality control methods in infection control and hospital epidemiology - Part I: introduction and basic theory. **Infection Control and Hospital Epidemiology**, v. 19, n. 4, p.194-214, 1998.

BENNEYAN, J. C. Statistical quality control methods in infection control and hospital epidemiology - Part II: chart use, statistics properties, and research issues. **Infection Control and Hospital Epidemiology**, v. 19, n. 4, p. 265-283, 1998.

BONITA, R.; BEAGLEHOLE, R.; KJELLSTRÖM, T. **Epidemiologia Básica**. 2.ed., São Paulo: Livraria Editora Santos, 2010. 213p.

BORTMAN, M. Elaboración de corredores o canales endémicos mediante planillas de cálculo. **Revista Panamericana de Salud Pública/Pan American Journal of Public Health**, v.5 n.1, p. 1-8, 1999.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. **Time series analysis**: forecasting and control, New Jersey: Prentice-Hall, 1994.

BRASIL. Diário Oficial da União. Lei nº 8080/90. Dispõe sobre as condições para promoção, proteção e recuperação da saúde, a organização e o financiamento dos serviços correspondentes e da outras providências. Brasília - DF, 19 de setembro de 1990.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia de Vigilância Epidemiológica. Brasília - DF: Ministério da Saúde, 2005

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Sinan [dados na Internet]. Brasília: MS. Disponível em: <<http://dtr2004.saude.gov.br/sinanweb/>>. Acesso em: 14 mar. 2012.

BRAZ RM. **Detecção precoce de epidemias de malária no Brasil**: uma proposta de automação. 122f. Dissertação (Mestrado em Saúde Pública) – Escola Nacional de Saúde Pública, Rio de Janeiro, 2005.

BROOKMEYER, R.; STROUP, D. **Monitoring the health of populations**: statistical principles & methods for public health surveillance. New York: Oxford University Press, 2004.

BURNS, C. M. et al. The use of CUSUM analysis in the early detection and management of hospital bed occupancy crises. **Medical Journal of Australia**, v. 183, n. 6, p. 291-294, 2005.

CASTILLO E. **Statistical Adjustment for Quality Control**, New York: John Wiley & Sons. Inc., 2002.

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. **Controle estatístico de qualidade**. São Paulo: Atlas, 2004.

COSTA, A. J. L. **Revisão das técnicas estatísticas propostas na literatura para detecção de epidemias de doenças infecciosas**. 253f. Dissertação (Mestrado em Saúde Coletiva) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 1994.

CULLEN, J. R. et al. An Epidemiological early warning system for malaria control in northern Thailand. **Bulletin of the World Health Organization**, v. 62, n. 1, p. 107-114, 1984.

CUSTÓDIO, C. A. **Gráficos de Controle CUSUM**: um enfoque dinâmico para a análise estatística de processos. 134f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

DAS, S. et al. Identification of Hot and Cold spots in genome of Mycobacterium tuberculosis using Shewhart Control Charts. **Scientific Reports**, v. 2, n. 297, p 1-6, 2012

DAVIS, M. M.; AQUILINO, N. J.; CHASE, R. B. **Fundamentos da administração da produção**. 3 ed. Porto Alegre: Bookman, 2001.

EWAN, W.D. When and how to use CUSUM charts. **Technometrics**, v.5, n.1, p.1-32, 1963.

FARRINGTON, P.; ANDREWS, N. Outbreak detection: application to infectious disease surveillance. In: STROUP, D. F.; BROOKMEYER, R. (Org.). **Monitoring the health of populations: statistical principles & methods for public health surveillance**, Oxford: Oxford University Press, 2004. p. 203-230

FREITAS, H. S. A.; MERCHÁN-HAMANN, E. Aspectos epidemiológicos da meningite pelo Haemophilus influenzae no Distrito Federal, Brasil, antes da introdução da vacina conjugada. **Epidemiologia e Serviços de Saúde**, v. 13, n. 1, p. 25 – 34, 2004.

FUNDAÇÃO NACIONAL DE SAÚDE. Guia de Vigilância Epidemiológica. 5 ed. Brasília: Ministério da Saúde, 1998.

GOMES, I. C.; MINGOTI, S. A.; OLIVEIRA, C. L. A novel experience in the use of control charts for the detection of nosocomial infection outbreaks. **Clinics** v. 66, n. 10, p. 1681-1689, 2011.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5. ed. Porto Alegre: Bookman, 2011.

HOOVER, R. et al. Cancer by county, new resources for etiologic clues. **Science**, v.189, p.1005-1007, 1975.

HUNTER, J. S. The exponentially weighted moving average. **Journal of Quality Technology**, v. 18, p. 203-210, 1986.

JEKEL, J. F. **Epidemiologia, Bioestatística e Medicina Preventiva**. Porto Alegre: Artes Médicas, 1999.

JOHNSON, N. L. A simple theoretical approach to cumulative sum control charts. **Journal of the American Statistical Association**, v. 56, p. 835 -840, 1961.

JURAN, J. M.; GRZYNA, F. M. Controle da Handbook. **Métodos Estatísticos Clássicos aplicados à Qualidade**, São Paulo: Makron Books, 1992.

KANE, V. E. Process Capability Indices. **Journal of Quality Technology**, v. 18, n. 1, p. 41-52, 1986.

KUNADIAN, B. et al. Cumulative funnel plots for the early detection of interoperator variation: retrospective database analysis of observed versus predicted results of percutaneous coronary intervention. **British Medical Journal**, v. 336, p. 931–934, 2008.

LIMA, S. S. Modelagem Estatística para o Monitoramento de Doenças de Notificação Compulsória. 86f. Dissertação (Mestrado em Matemática e Estatística) – Universidade Federal do Pará, Belém, 2008.

LUCAS, M. J.; SACCUCCI, M. S. Exponentially weighted moving average control schemes: properties and enhancements. **Technometrics**, v. 32, n. 1, p. 1-12, 1990.

MOHAMMED, M. A. et al. A practical method for monitoring general practice mortality in the UK: findings from a pilot study in a health board of Northern Ireland British **British Journal of General Practice**, v. 55, p. 670–676, 2005.

MONTGOMERY, D. C. **Introdução ao controle estatístico da qualidade**. Tradução de Ana Maria Lima de Farias. 4 ed. Rio de Janeiro: LTC, 2004.

MONTGOMERY, D. C. **Introduction to statistical process control**. 3 ed., New York: John Wiley & Sons, Inc., 1997.

MONTOYA-RESTREPO, N. E.; CORREA-MORALES, J. C. Cartas de Control Estadístico de Procesos en el Monitoreo de la Mortalidad Perinatal. **Revista de Salud Pública**, v. 11, n. 1, p. 92-99, 2009.

NOYEZ, L. Control charts, CUSUM techniques and funnel plots. A review of methods for monitoring performance in healthcare. **Interactive Cardiovascular and Thoracic Surgery**, v. 9, p. 494–499, 2009.

PAGE, E. S. Continuous inspection schemes. **Biometrika**, v. 41, p. 100–115, 1954.

PEREIRA, Z. L.; REQUEIJO, J. G. **Qualidade: Planejamento e Controle Estatístico de Processo**. Lisboa: Prefácio, 2008.

PINDYCK, R. S.; RUBINFELD, D. L. **Econometria**. Rio de Janeiro: Elsevier, 2004.

PORTA, M.; GREENLAND, S.; LAST, J.M. **A dictionary of epidemiology**. 5 ed. New York: Oxford University Press, 2008.

PYZDEK, T.; KELLER, P. A. **Seis sigma**: guia do profissional, uma guia completo para Green Belts, Black Belts e gerentes em todos os níveis. 3 ed. Rio de Janeiro, RJ: Alta Books, 2011.

RICH, W. H; TERRY, M. C. The industrial control chart applied to the study of epidemics. **Public Health Reports**, v. 61, n 42, p. 1501 – 1511, 1946.

ROGERS C. A. et al. Control chart methods for monitoring cardiac surgical performance and their interpretation. **Journal of Thoracic and Cardiovascular Surgery**, v.128, p. 811-819, 2004.

ROSSI, G.; LAMPUGNANI, L.; MARCHI, M. An approximate CUSUM procedure for surveillance of health events. **Statistics in Medicine**, v. 18, p. 2111-2122, 1999.

SANCHES, O. Princípios básicos de procedimentos estatísticos aplicados na análise de dados de vigilância em saúde pública: uma revisão. **Caderno de Saúde Pública**, Rio de Janeiro, v. 16, n. 2, p. 317-333, abr./jun. 2000.

SCHILLING, W. Analysis of the Data of a Public Health Organization by the Control Chart Method. **Journal of the American Statistical Association**, v. 39, n. 227, p. 311-324, set. 1944.

SCHRAMM, J. M. A et al. Transição epidemiológica e o estudo de carga de doença no Brasil. **Ciência e Saúde Coletiva**, Rio de Janeiro, v. 9, n. 4, p. 897-908, 2004.

SECRETARIA DE VIGILÂNCIA EM SAÚDE. Guia de vigilância epidemiológica. 6ª ed. Brasília: Ministério da Saúde; 2005.

SEGO, L. H.; WOODALL, W. H.; REYNOLDS JUNIOR, M. R. A comparison of surveillance methods for small incidence rates. **Statistics in Medicine**, v. 27, p. 1225–1247, 2008.

SELLICK, J.A. The use of statistical process control charts in hospital epidemiology. *Infection Control and Hospital Epidemiology*, v. 14, n. 11, p. 649-656, 1993.

SERFLING, W. Analysis of the data of public health organization by the control chart method. **Journal of the American Statistical Association**, v. 39, p. 311-324, 1944.

SHEWHART, W. A. **Economic Control of Quality of Manufactured Product**. New York: D.Van Nostrand Co. Inc., 1931.

SHU, L.; JIANG, W.; TSUI, K. L. A comparison of weighted CUSUM procedures that account for monotone changes in population size. **Statistics in Medicine**, v. 30, p. 725–741, 2011.

VASCONCELLOS, M. A. S.; ALVES, D. (Orgs.). **Manual de Econometria**. 1 ed. São Paulo: Atlas, 2000.

WERKEMA, M. C. C. **Ferramentas estatísticas básicas para o gerenciamento de processos**. Belo Horizonte: QFCO, 1995.

WILLIAMSON, G. D.; HUDSON, G. W. A monitoring system for detecting aberrations in public health surveillance reports. **Statistics in Medicine**, v. 18, p. 3283-3298, 1999.

WOODALL, W. H. On the use and evaluation of prospective scan methods for health-related surveillance. **Journal of the Royal Statistical Society**. v. 171, p. 223-237, 2008.

WOODALL, W. H. Use of control charts in health care and public health surveillance (with discussion). **Journal of Quality Technology**, v. 38, p. 88-103, 2006.

ZANINI, R. R. **A utilização de gráficos de controle: uma aplicação na área da saúde**. 2006. 110f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Maria, Santa Maria, 2006.

ANEXOS

Anexo A - Fatores para o cálculo dos limites de controle (3σ) para o gráfico de médias.

Número de Elementos na Amostra (n)	Fatores para Limites de Controle		
	A ₁	A ₂	A ₃
2	2,121	1,880	2,659
3	1,732	1,023	1,954
4	1,5	0,729	1,628
5	1,342	0,577	1,427
6	1,225	0,483	1,287
7	1,134	0,419	1,182
8	1,061	0,373	1,099
9	1	0,337	1,032
10	0,949	0,308	0,975
11	0,905	0,285	0,927
12	0,866	0,266	0,886
13	0,832	0,249	0,850
14	0,802	0,235	0,817
15	0,775	0,223	0,789

Anexo B - Fatores para o cálculo dos limites de controle (3σ) para o gráfico do desvio padrão.

Número de Elementos na Amostra (n)	Fatores para Linha Central		Fatores para Limites de Controle			
	c_4	$1/c_4$	B_3	B_4	B_5	B_6
2	0,7979	1,2533	0	3,267	0	2,606
3	0,8862	1,1284	0	2,568	0	2,276
4	0,9213	1,0854	0	2,266	0	2,088
5	0,9400	1,0638	0	2,089	0	1,964
6	0,9515	1,0510	0,030	1,970	0,029	1,874
7	0,9594	1,0423	0,118	1,882	0,113	1,806
8	0,9650	1,0363	0,185	1,815	0,179	1,751
9	0,9693	1,0317	0,239	1,761	0,232	1,707
10	0,9727	1,0281	0,284	1,716	0,276	1,669
11	0,9754	1,0252	0,321	1,679	0,313	1,637
12	0,9776	1,0229	0,354	1,646	0,346	1,610
13	0,9794	1,0210	0,382	1,618	0,374	1,585
14	0,9810	1,0194	0,406	1,594	0,399	1,563
15	0,9823	1,0180	0,428	1,572	0,421	1,544

Anexo C - Fatores para o cálculo dos limites de controle (3σ) para o gráfico da amplitude.

Número de Elementos na Amostra (n)	Fatores para Linha Central			Fatores para Limites de Controle			
	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
2	1,128	0,8865	0,853	0	3,686	0	3,267
3	1,693	0,5907	0,888	0	4,358	0	2,574
4	2,059	0,4857	0,880	0	4,698	0	2,282
5	2,326	0,4299	0,864	0	4,918	0	2,114
6	2,534	0,3946	0,848	0	5,078	0	2,004
7	2,704	0,3698	0,833	0,204	5,204	0,076	1,924
8	2,847	0,3512	0,820	0,388	5,306	0,136	1,864
9	2,970	0,3367	0,808	0,547	5,393	0,184	1,816
10	3,078	0,3249	0,797	0,687	5,469	0,223	1,777
11	3,173	0,3152	0,787	0,811	5,535	0,256	1,744
12	3,258	0,3069	0,778	0,922	5,594	0,283	1,717
13	3,336	0,2998	0,770	1,025	5,647	0,307	1,693
14	3,407	0,2935	0,763	1,118	5,696	0,328	1,672
15	3,472	0,2880	0,756	1,203	5,741	0,347	1,653

Anexo D - Modelos selecionados para os casos de Hepatite C, tamanho amostral válido, média, valores mínimo e máximo, desvio padrão, variância e valores dos critérios AIC e BIC.

	N	Média	Mínimo	Máximo	Desvio Padrão	Variância	AIC	BIC
(1,1,0)(0,0,0) ₁₂	59	1,18	-99,06	122,14	40,64	1651,30	7,44	7,48
(1,0,0)(0,0,0) ₁₂	60	7,83	-109,33	156,00	49,90	2489,66	7,85	7,89
(1,1,0)(1,0,0) ₁₂	59	0,72	-91,16	110,16	38,36	1471,76	7,33	7,36
(0,1,1)(1,0,0)₁₂	59	2,16	-59,55	107,04	35,37	1250,97	7,17	7,20
(1,0,0)(1,0,0) ₁₂	60	8,83	-128,67	156,00	46,21	2134,96	7,70	7,73
(2,0,0)(1,0,0) ₁₂	60	6,20	-80,59	156,00	42,87	1838,02	7,58	7,65
(2,1,0)(1,0,0) ₁₂	59	1,13	-88,04	115,28	35,74	1277,37	7,22	7,29
(2,1,0)(1,1,0) ₁₂	58	-0,54	-138,39	127,72	46,20	2134,33	7,73	7,81
(1,0,0)(1,1,0) ₁₂	59	0,72	-91,16	110,16	38,36	1471,76	7,33	7,36
(2,0,0)(1,1,0) ₁₂	59	1,13	-88,04	115,28	35,74	1277,37	7,22	7,29
(2,0,0)(1,1,0) ₁₂	58	-0,54	-138,39	127,72	46,20	2134,33	7,73	7,81

Anexo E - Somas acumulativas dos dados de hepatite C,. Rio Grande do Sul, 2007 a 2011.

(continua)

Período	X_i	$x_i-(m_1+k)$	C^+	N^+	$(m_1-k)-x_i$	C^-	N^-
2	-1,00	-3,16	0,00	0	14,34	0,00	0
3	11,19	9,03	0,00	0	26,53	0,00	0
4	-11,98	-14,14	0,00	0	3,36	0,00	0
5	36,35	34,18	0,00	0	51,68	0,00	0
6	25,29	23,12	0,00	0	40,62	0,00	0
7	-1,62	-3,79	0,00	0	13,71	0,00	0
8	-0,31	-2,47	0,00	0	15,03	0,00	0
9	-32,25	-34,41	0,00	0	-16,91	16,91	1
10	10,01	7,85	0,00	0	25,35	0,00	0
11	-39,93	-42,09	0,00	0	-24,59	24,59	1
12	-48,18	-50,34	0,00	0	-32,84	57,43	2
13	42,18	40,02	4,65	1	57,52	0,00	0
14	-21,55	-23,71	0,00	0	-6,21	6,21	1
15	24,08	21,91	0,00	0	39,41	0,00	0
16	38,13	35,97	0,60	1	53,47	0,00	0
17	16,41	14,25	0,00	0	31,75	0,00	0
18	95,08	92,91	57,55	1	110,41	0,00	0
19	-3,19	-5,36	0,00	0	12,14	0,00	0
20	17,96	15,80	0,00	0	33,30	0,00	0
21	-2,70	-4,86	0,00	0	12,64	0,00	0
22	-34,85	-37,02	0,00	0	-19,52	19,52	1
23	10,15	7,99	0,00	0	25,49	0,00	0
24	15,60	13,43	0,00	0	30,93	0,00	0
25	-37,96	-40,13	0,00	0	-22,63	22,63	1
26	-3,64	-5,81	0,00	0	11,69	10,93	2
27	51,29	49,12	13,76	1	66,62	0,00	0
28	19,16	16,99	0,00	0	34,49	0,00	0
29	53,19	51,03	15,66	1	68,53	0,00	0
30	65,80	63,63	28,26	2	81,13	0,00	0
31	9,72	7,55	0,00	0	25,05	0,00	0
32	-57,90	-60,06	0,00	0	-42,56	42,56	1
33	3,17	1,01	0,00	0	18,51	24,06	2
34	26,97	24,81	0,00	0	42,31	0,00	0
35	-26,68	-28,85	0,00	0	-11,35	11,35	1
36	-8,50	-10,66	0,00	0	6,84	4,51	2
37	6,17	4,01	0,00	0	21,51	0,00	0
38	-34,49	-36,65	0,00	0	-19,15	19,15	1
39	21,00	18,83	0,00	0	36,33	0,00	0
40	-29,74	-31,90	0,00	0	-14,40	14,40	1
41	107,04	104,88	69,51	1	122,38	0,00	0

(conclusão)

Período	X_i	$x_i-(m_1+k)$	C^+	N^+	$(m_1-k)-x_i$	C^-	N^-
42	-59,55	-61,71	0,00	0	-44,21	44,21	1
43	-27,60	-29,76	0,00	0	-12,26	56,48	2
44	7,71	5,54	0,00	0	23,04	33,44	3
45	-11,01	-13,17	0,00	0	4,33	29,11	4
46	-29,75	-31,91	0,00	0	-14,41	43,52	5
47	-17,97	-20,14	0,00	0	-2,64	46,15	6
48	6,49	4,33	0,00	0	21,83	24,32	7
49	-16,01	-18,17	0,00	0	-0,67	24,99	8
50	-22,83	-24,99	0,00	0	-7,49	32,49	9
51	-38,32	-40,48	0,00	0	-22,98	55,47	10
52	-21,39	-23,56	0,00	0	-6,06	61,52	11
53	-40,74	-42,90	0,00	0	-25,40	86,92	12
54	-38,15	-40,32	0,00	0	-22,82	109,74	13
55	18,06	15,90	0,00	0	33,40	76,34	14
56	51,70	49,54	14,17	1	67,04	9,31	15
57	6,12	3,96	0,00	0	21,46	0,00	0
58	37,96	35,79	0,42	1	53,29	0,00	0
59	47,61	45,44	10,08	2	62,94	0,00	0
60	-34,15	-36,31	0,00	0	-18,81	18,81	1

Anexo F - Lista Nacional de Doenças e Agravos de Notificação Compulsória, Secretaria de Vigilância em Saúde /MS, 2002.

Doenças e Agravos	
Botulismo	Febre maculosa
Carbúnculo ou “antraz”	Raiva humana
Cólera	Rubéola
Coqueluche	Síndrome da rubéola congênita
Leishmaniose tegumentar americana	Sarampo Febre tifóide
Leishmaniose visceral	Hanseníase
Leptospirose	Hantavíruses
Malária	Hepatites virais
Dengue	Sífilis congênita
Difteria	Síndrome da imunodeficiência adquirida (aids)
Doença de Chagas (casos agudos)	Síndrome respiratória aguda grave
Doenças meningocócicas e outras meningites	Tétano
Meningite por <i>Haemophilus influenzae</i>	Infecção pelo vírus da imunodeficiência humana (HIV)
Peste	em gestantes e crianças expostas ao risco de transmissão vertical
Poliomielite	Tularemia
Paralisia flácida aguda	Tuberculose
Esquistossomose (em área não-endêmica)	Variola
Febre amarela	
Febre do Nilo	