

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO**

Elisa Henning

**APERFEIÇOAMENTO E DESENVOLVIMENTO DOS
GRÁFICOS COMBINADOS SHEWHART-CUSUM BINOMIAIS**

Tese submetida ao Programa de Engenharia de Produção da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Engenharia de Produção.
Orientador: Prof. Robert Wayne Samohyl, PhD.

Florianópolis

2010

Catálogo na fonte elaborada pela biblioteca da
Universidade Federal de Santa Catarina

H517a Henning, Elisa

Aperfeiçoamento e desenvolvimento dos gráficos combinados
Shewhart-Cusum binomiais [tese] / Elisa Henning ;
orientador, Robert Wayne Samohyl. - Florianópolis, SC, 2010.
251 p. : tabs., grafs.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Engenharia de Produção.

Inclui referências

1. Engenharia de produção. 2. Engenharia - Gráficos.
I. Samohyl, Robert Wayne. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Engenharia de Produção.
III. Título.

CDU 658.5

Elisa Henning

**APERFEIÇOAMENTO E DESENVOLVIMENTO DOS
GRÁFICOS COMBINADOS SHEWHART-CUSUM BINOMIAIS**

Esta Tese foi julgada adequada para obtenção do Título de “Doutor em Engenharia de Produção”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção.

Florianópolis, 12 de Novembro de 2010.

Prof. Antônio Cezar Bornia, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Robert Wayne Samohyl,
Ph.D.
Orientador

Prof^a. Sueli Fischer Beckert, Dra.
Instituto Superior Tupy

Prof. Edson Marcos Leal Soares
Ramos, Dr.
Universidade Federal do Pará

Prof. Marcelo Menezes Reis, Dr.
Universidade Federal de Santa
Catarina

Prof^a. Fernanda Cristina Barbosa
Pereira Queiroz, Dra.
Universidade Federal do Rio
Grande do Norte

Prof^a. Vera do Carmo Comparsi de
Vargas, Dra.
Universidade Federal de Santa
Catarina

Para Aninha e Joaquim, sem eles a
vida não tem sentido...

AGRADECIMENTOS

Ao professor orientador, Robert Wayne Samohyl, uma pessoa fascinante, pela colaboração, incentivo, dedicação e paciência em todas as etapas da execução deste trabalho.

A todos os professores, funcionários (especialmente à Meri) e colegas do Departamento de Engenharia de Produção da UFSC.

À UDESC - Universidade do Estado de Santa Catarina pela licença e suporte financeiro concedidos.

À UNIVILLE – Universidade da Região de Joinville, pela licença cedida.

Aos colegas do NNQ - Núcleo de Normalização e Qualimetria, Rodrigo, Gueibi, Eder e Rubson – foi um prazer conhecê-los.

À Andrea, Custódio e Olga, que muito me ajudaram, trocando artigos, livros, idéias e desabafos.

A todos meus colegas do Departamento de Matemática do CCT-UDESC pela colaboração no período em que estive afastada, em especial a Lígia, Marnei, Ivanete e Rogério, pelo incentivo e companheirismo.

À engenheira Helena Dausacker da Cunha, da Companhia Águas de Joinville, e ao professor Leandro Zvirtes, do Departamento de Engenharia de Produção da UDESC, por disponibilizarem dados reais para análise neste trabalho.

À minha família, pela compreensão, e por suportar as minhas longas ausências, nem sempre só físicas.

Ao meu pai Martin (em memória) que sempre incentivou todos ao seu redor, inclusive eu, a continuar estudando.

A todos aqueles que, de algum modo, contribuíram para este trabalho, meu sincero reconhecimento e gratidão.

Enfim, agradeço a Deus, graças a Ele, estou aqui...

Quando não está em nosso poder seguir o que é verdade, seguimos o que é mais provável.

René Descartes

RESUMO

Os tradicionais gráficos de controle Shewhart são considerados efetivos na detecção de grandes mudanças na média, variância ou na fração não conforme, enquanto que gráficos de controle de soma cumulativa (CUSUM) são recomendados para a sinalização de pequenas e moderadas alterações nestes parâmetros. Nenhum dos gráficos mencionados terá um bom desempenho em todas as situações. Uma solução possível para este problema é combinar múltiplos gráficos para abranger mudanças de diversas magnitudes. Assim, um gráfico combinado Shewhart-CUSUM tem como finalidade aumentar a sensibilidade do procedimento CUSUM para alterações maiores. Este trabalho traz várias contribuições para o desenvolvimento e aperfeiçoamento de gráficos combinados Shewhart-CUSUM para dados com distribuição binomial. Inicialmente, a partir do resultado de simulações, analisa-se o desempenho de um gráfico combinado e, se a adição de linhas Shewhart a um gráfico CUSUM binomial unilateral superior realmente aumenta a sensibilidade deste. O desempenho de um gráfico combinado Shewhart-CUSUM é também comparado com o gráfico tipo Shewhart e com procedimentos CUSUM delineados para detecção de mudanças maiores. Pensando em aplicações, foi elaborada uma metodologia para construção de um gráfico combinado incluindo a análise das suposições necessárias (aderência, autocorrelação e superdispersão). Para finalizar, esta metodologia foi aplicada a dados adaptados da literatura e também de processos reais. O trabalho ainda contempla algumas contribuições adicionais como o uso de limites exatos (ou probabilísticos) na parte Shewhart do gráfico combinado e uma proposta de aproximação para o limite superior do CUSUM binomial. Os resultados obtidos revelam que o gráfico combinado Shewhart-CUSUM aumenta a sensibilidade de um gráfico CUSUM binomial para magnitudes de mudança maiores que as de planejamento e identificou-se a existência de uma região onde o gráfico combinado tem desempenho superior aos dois gráficos individuais. Os resultados das aplicações foram satisfatórios, validando a metodologia elaborada. A partir das aplicações foram sinalizadas situações práticas onde o gráfico combinado é mais efetivo que os gráficos individuais.

Palavras-chave: Gráfico Shewhart “np”. Gráfico CUSUM binomial. Gráfico Combinado Shewhart-CUSUM binomial. Comparação de desempenho.

ABSTRACT

The traditional Shewhart control charts are considered effective in detecting large changes in mean, variance or the fraction nonconforming, while cumulative sum control charts (CUSUM) are recommended for recognizing small and moderate changes in these parameters. None of the graphics mentioned will perform well in all situations. One possible solution to this problem is to combine multiple charts to cover changes of various magnitudes. Thus, a combined Shewhart-CUSUM chart is intended to increase the sensitivity of the CUSUM procedure to large shifts. This work brings several contributions to the development and improvement of combined Shewhart-CUSUM charts for data with the binomial distribution. Initially, from the results of simulations, we analyze the performance of the combined scheme and if the addition of the Shewhart chart lines in a binomial CUSUM actually increases the sensitivity of this chart. The performance of the combined Shewhart-CUSUM chart is also compared with the Shewhart chart and CUSUM procedures designed to detect major changes. Thinking of applications, a methodology was developed for building a combined scheme including analysis of the assumptions required (goodness of fit, autocorrelation and over dispersion). Finally the methodology is applied to data adapted from literature and also to real processes. The work also includes some additional contributions, as the use of exact probabilistic limits and a proposed approach to the upper limit of the binomial CUSUM. The results show that the combined Shewhart-CUSUM chart is more sensitive than a binomial CUSUM chart for larger than planned changes, and we identified the existence of a region where the combined chart has better performance than the CUSUM chart and also to Shewhart-type chart. The results of the applications were satisfactory, thus validating the methodology developed. From several applications, practical situations were flagged where the combined chart is more effective than the individual charts.

Keywords: Shewhart “np” chart. Binomial CUSUM chart. Combined Shewhart-CUSUM schemes for binomial data. Comparison of performance.

LISTA DE FIGURAS

Figura 2. 1 – Gráfico de Controle do tipo Shewhart.	41
Figura 2. 2 – Probabilidade de ocorrência de observações além de três desvios padrão da média em uma distribuição normal.	47
Figura 3. 1 – Exemplo de um CUSUM bilateral.....	58
Figura 3. 2 – Exemplo de um CUSUM unilateral superior.....	58
Figura 4. 1 – Exemplo de gráfico combinado bilateral.....	84
Figura 4. 2 – Exemplo de gráfico combinado unilateral superior.....	84
Figura 4. 3 – Comparação de valores de $\log(ARL)$ da Tabela 4.1.....	86
Figura 5. 1 – Comparação de valores de h : raiz ($h.cd$) e das aproximações h (2 termos), 3 termos (h_1 - Equação 5.12), 4 termos (h_2 -Equação 5.15) e 5 termos (h_3 -Equação 5.16).....	102
Figura 5. 2 – Gráfico comparativo dos valores de h (Tabela 5.2).....	104
Figura 5. 3 – Fluxograma do processo de simulação para ARL.....	106
Figura 5. 4 – Fluxograma do processo de aproximação de h	111
Figura 6. 1 – Histogramas do valor de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 para $p = p_0$ (0,06).....	120
Figura 6. 2 – Gráfico violin plot dos valores de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 para $p = p_0$ (0,06).	121
Figura 6. 3 – Histogramas dos valores de ARL do procedimento combinado e da parte CUSUM para o Exemplo 6.1 com $p = p_1 = 0,066$	122
Figura 6. 4 – Gráfico violin plot dos valores de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 com $p = p_1 = 0,066$	122
Figura 6. 5 – Histogramas dos valores de ARL do gráfico combinado e somente parte CUSUM para o Exemplo 6.1 com $p = 0,09$	123
Figura 6. 6 – Histograma dos valores de ARL somente da parte Shewart do gráfico combinado para o Exemplo 6.1 com $p = 0,09$	124
Figura 6. 7 – Gráfico violin plot dos valores de ARL do gráfico combinado Shewart-CUSUM e somente da parte CUSUM para o Exemplo 6.1 com	124
Figura 6. 8 – Gráfico Boxplot dos valores de ARL do gráfico combinado Shewart-CUSUM e das partes individuais (Shewart e CUSUM) para o Exemplo 6.1 com $p = 0,09$	125
Figura 6. 9 – Histogramas dos valores de ARL do gráfico combinado Shewart-CUSUM e das partes individuais (Shewart e CUSUM) para o Exemplo 6.1 com $p = 0,102$	126
Figura 6. 10 – Gráfico violin plot dos valores de ARL do gráfico combinado Shewart-CUSUM e das partes individuais correspondentes para o Exemplo 6.1 com $p = 0,102$	126
Figura 6. 11 – Histogramas dos valores de ARL do gráfico combinado Shewart-CUSUM e das partes individuais (Shewart e CUSUM) para o Exemplo 6.1 com $p = 0,114$	127
Figura 6. 12 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewart-CUSUM e das partes individuais (Shewart e CUSUM) para o Exemplo 6.1 com $p = 0,114$	127

Figura 6. 13 – Histograma dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,120$.	128
Figura 6. 14 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,120$.	128
Figura 6. 15 – Histogramas dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,150$.	129
Figura 6. 16 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,150$.	129
Figura 6. 17 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,240$.	130
Figura 6. 18 – Gráficos dos valores de ARL do gráfico combinado Shewhart-CUSUM e da parte CUSUM individual para o Exemplo 6.2 com $p = 0,02$; 0,025; 0,04 e 0,05.	132
Figura 6. 19 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais Shewhart e CUSUM para o Exemplo 6.2 com $p = 0,065$; 0,08; 0,10 e 0,15.	132
Figura 6. 20 – Comparação dos valores de $\log(\text{ARL})$ para o exemplo 6.3.	136
Figura 6. 21 – Comparação dos valores de $\log(\text{MRL})$ para o exemplo 6.3.	136
Figura 6. 22 – $r\text{ARL}$ para o Exemplo 6.3.	137
Figura 6. 23 – Comparação de SDRL e variação do SDRL, para o Exemplo 6.3.	138
Figura 6. 24 – Comparação de $\log(\text{ARL})$, para o Exemplo 6.4.	139
Figura 6. 25 – Região onde o gráfico combinado é mais efetivo que os gráficos CUSUM e tipo Shewhart para o Exemplo 6.4.	141
Figura 6. 26 - Comparação de $\log(\text{MRL})$, para o Exemplo 6.4.	141
Figura 6. 27 - Comparação de $\log(\text{SDRL})$, para o Exemplo 6.4.	142
Figura 6. 28 - $r\text{ARL}$ (CUSUM e combinado), para o Exemplo 6.4.	143
Figura 6. 29 - $r\text{ARL}$ (Combinado e Shewhart), para o Exemplo 6.4.	143
Figura 6. 30 – $r\text{ARL}$ entre os gráficos CUSUM e tipo Shewhart em relação ao gráfico combinado para o Exemplo 6.4.	144
Figura 6. 31 – Redução do ARL do gráfico combinado, para o Exemplo 6.4.	145
Figura 6. 32 – Comparação de valores $\log(\text{ARL})$, para o Exemplo 6.5.	147
Figura 6. 33 – Região onde o gráfico combinado é mais efetivo que os gráficos CUSUM e tipo Shewhart para o Exemplo 6.4.	147
Figura 6. 34 – $r\text{ARL}$ (Combinado e CUSUM) para o exemplo 6.5.	148
Figura 6. 35 – $r\text{ARL}$ (Combinado e Shewhart), para o Exemplo 6.5.	148
Figura 6. 36 – Redução de ARL, para o Exemplo 6.5.	149
Figura 7. 1 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM, para o Exemplo 7.1.	155

Figura 7. 2 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 1.5, para o Exemplo 7.1.	155
Figura 7. 3 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 2.0, para o Exemplo 7.1.	156
Figura 7. 4 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 3.0, para o Exemplo 7.1.	156
Figura 7. 5 – Comparação de $\log(\text{ARL})$ dos Gráficos Combinado e procedimentos CUSUM, para o Exemplo 7.1.	157
Figura 7. 6 – Comparação de $r\text{ARL}$ para os gráficos combinado, CUSUM e CUSUM 1.5, do Exemplo 7.1.	157
Figura 7. 7 – Comparação de $r\text{ARL}$ para os gráficos combinado e CUSUM 2.0 e 3.0, do Exemplo 7.1.	158
Figura 7. 8 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 1.1e 1.5, para o Exemplo 7.2.	159
Figura 7. 9 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado, CUSUM 2.0, e Shewhart para o Exemplo 7.2.	160
Figura 7. 10 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM, CUSUM 1.5 e CUSUM 2.0 para o Exemplo 7.2.	161
Figura 7. 11 – Saída do programa anygeth para ARL sem FIR.	162
Figura 7. 12 – Saída do programa anygeth para ARL com FIR.	162
Figura 7. 13 – Saída do R para ARL sem e com FIR.	163
Figura 8. 1 – Fluxograma para construção de um gráfico combinado.	166
Figura 8. 2 – Gráfico combinado para os dados de xícaras defeituosas.	169
Figura 8. 3 – Gráfico CUSUM para os dados de xícaras defeituosas.	170
Figura 8. 4 – Gráfico tipo Shewhart para os dados de xícaras defeituosas.	170
Figura 8. 5 – Gráficos da função de autocorrelação e autocorrelação parcial para os dados de coliformes totais.	172
Figura 8. 6 – Gráfico Combinado Shewhart-CUSUM para os dados de Coliformes Totais.	174
Figura 8. 7 – Gráfico Shewhart para os dados de Coliformes Totais.	175
Figura 8. 8 – Gráfico CUSUM para os dados de Coliformes Totais $p_1 = 0,05$	175
Figura 8. 9 – Gráfico Shewhart para os dados de Coliformes Totais com regras suplementares.	176
Figura 8. 10 – Gráfico Combinado Shewhart-CUSUM com $ARL_0 = 465,5e$ limites ajustados.	176
Figura 8. 11 – Gráficos da função de autocorrelação e autocorrelação parcial para os dados da rotulagem de garrafas.	179
Figura 8. 12 – Gráfico Shewhart-CUSUM para os dados da rotulagem de garrafas.	180
Figura 8. 13 – Gráfico CUSUM para os dados da rotulagem de garrafas.	181
Figura 8. 14 – Gráfico CUSUM para os dados da rotulagem de garrafas.	181
Figura 8. 15 – Gráfico Shewhart para os dados da rotulagem de garrafas.	182
Figura 8. 16 – Gráfico Shewhart com regras sensibilizantes para os dados da rotulagem de garrafas.	183

Figura I. 1 – Teste de aderência para dados do Exemplo 8.2.....	232
Figura I. 2 – Teste de aderência para dados do Exemplo 8.3.....	233
Figura E. 1: violin plots para dados simulados com distribuição bimodal ($\mu_1=-2$, $\mu_2=2$ e $\sigma_1=\sigma_2=0,6$), uniforme ($\alpha = -4$ e $\beta = 4$) e normal ($\mu=0$ e $\sigma=3$).....	246
Figura E. 2: Boxplots para dados simulados com distribuição bimodal ($\mu_1 =$ -2 , $\mu_2 = 2$ e $\sigma_1 = \sigma_2 = 0,6$), uniforme ($\alpha = -4$ e $\beta = 4$) e normal ($\mu =$ 0 e $\sigma = 3$).....	247

LISTA DE TABELAS

Tabela 2. 1 –Valores calculados de ARL_0	48
Tabela 3. 1 – Comparação entre gráficos de controle Shewhart e CUSUM.....	82
Tabela 4. 1 - Valores de ARL para os seguintes gráficos: CUSUM ($k = 0,5$ e $h = 5$); combinado ($k = 0,5$; $h = 5,1$ e $LSC = 4,0$) e combinado ($k = 0,5$; $h = 5,2$ e $LSC = 3,5$).....	86
Tabela 5. 1 – Valores mínimos de qc para considerando s (acréscimo em p_0) e erro relativo de 5% e 2%.....	101
Tabela 5. 2 – Valores de h aproximados para comparação	103
Tabela 5. 3– Valores de ARL_0 com algoritmo de Hawkins (ARL_{C_0})e simulados (ARL_{S_0}) para h (aproximado com a Equação 5.15) e k (calculado pela Equação 3.11) para um ARL_0 alvo de 300).....	106
Tabela 5. 4– Valores de ARL_0 com algoritmo de Hawkins (ARL_{C_0})e simulados (ARL_{S_0}) com h (aproximado com a Equação 5.15) e k (calculado pela Equação 3.11) para ARL_0 alvo de 370, $n=100$ e $p_0=0,05$	107
Tabela 5. 5 - Valores de ARL_1 simulados (ARL_{C_1}) e calculados (ARL_{S_1}) para $p_0 = 0,05$, p_1 , n e $ARL_0 = 370$	108
Tabela 5. 6 - Valores de h calculados pelo programa hrs.5, ARL_0 (ARL_{C_0}) calculado por cadeias de Markov e erro relativo (%), para.....	112
Tabela 5. 7 - Valores de h calculados pelo programa hrs.5, ARL_0 (ARL_{C_0}) calculado por Markov e erro relativo (%), com $p_0 > 0,1$;.....	114
Tabela 5. 8 - Valores de h calculados pelo programa hrs.5, ARL_0 calculado por cadeias de Markov e erro relativo (%), para $p_0 \leq 0,01$; $s = 25\%$, ARL_0 variado e tamanho de amostra $n = 100$ e 200	115
Tabela 6. 1- ARL, CSRL e CKRL para o gráfico combinado do Exemplo 6.1 e valores de ARL das partes CUSUM e Shewhart.....	120
Tabela 6. 2 – Valores de ARL, CSRL e CKRL para o gráfico combinado do Exemplo 6.2 e valores de ARL das partes CUSUM e Shewhart	131
Tabela 6. 3 - Valores de ARL, quartis e SDRL para os gráficos combinado (CS) e CUSUM (C) do exemplo 6.3	135
Tabela 6. 4 - ARL, MRL e SDRL, para o exemplo 6.4.	140
Tabela 6. 5 - ARL, MRL e SDRL, para o Exemplo 6.5.....	146
Tabela 6. 6 - Parâmetros de um gráfico combinado com.....	151
Tabela 6. 7 - Parâmetros de um gráfico combinado com $n = 50$; $s = 20\%$ e $ARL_0 = 370$	152
Tabela 7. 1 - Valores de ARL para o Exemplo 1.....	154
Tabela 7. 2 - Valores de ARL para o Exemplo 2.....	159
Tabela 8. 1 - Número de xícaras defeituosas por amostra	168
Tabela I. 1 - Valores observados e esperados para o exemplo 8.2.	231
Tabela I. 2 -Valores observados e esperados para os dados do exemplo 8.3...	232
Tabela F. 1 - Conjunto de dados do exemplo de aplicação 8.2: Número de amostras não conformes (com presença de coliformes totais).....	248

Tabela G. 1 - Conjunto de dados do exemplo 8.3: Processo de rotulagem de garrafas.....	249
--	-----

LISTA DE ABREVIATURAS E SIGLAS

- ACF - Função de autocorrelação amostral
- ARL - número médio de amostras coletadas até a emissão de um sinal (*Average Run Length*)
- ARL₀ - número médio de amostras coletadas até a emissão de um sinal durante o período sob controle
- ARL₁ - número médio de amostras coletadas até a emissão de um sinal que indique uma situação de fora de controle estatístico
- CD - Correção de difusão (*Corrected diffusion*)
- CEP - Controle Estatístico de Processos
- CUSUM - Soma acumulada (Cumulative Sum)
- CKRL - coeficiente de curtose do número de amostras coletadas até à emissão de um sinal (*Coefficient of Kurtosis Run Length*)
- CSRL - coeficiente de assimetria do número de amostras coletadas até à emissão de um sinal (*Coefficient of Skewness Run Length*)
- LSC - Limite superior do gráfico Shewhart
- MRL - mediana do número de amostras coletadas até à emissão de um sinal (*Median Run Length*)
- PACF - Função de autocorrelação amostral parcial
- RL - número de amostras coletadas até à emissão de um sinal (*Run Length*)
- RIR - Resposta Inicial Rápida (*Fast Initial Response*)
- SPRT - Teste Seqüencial da Razão de Probabilidades (Sequential Probability Ratio Test)
- SDRL - desvio padrão do número de amostras coletadas até à emissão de um sinal (*Standard Deviation Run Length*)

LISTA DE SÍMBOLOS

H_0 - hipótese nula

H_1 - hipótese alternativa

k - valor de referência do CUSUM

h - limite de controle superior do CUSUM

α - probabilidade de um erro tipo I

β - probabilidade de um erro tipo II

n - tamanho da amostra

p_0 - proporção considerada sob controle estatístico

p_1 - proporção fora de controle estatístico que se deseja detectar

q_1 - quartil inferior ou primeiro quartil

q_2 - segundo quartil ou mediana

q_3 - quartil superior ou terceiro quartil

s - alteração em p_0 (em termos de percentagem)

d - alteração em p_0 (em termos de proporção)

σ - desvio padrão

SUMÁRIO

1. INTRODUÇÃO	29
1.1 JUSTIFICATIVA	31
1.2 OBJETIVOS PROPOSTOS	32
1.3 MÉTODO	33
1.4 LIMITAÇÕES	35
1.5 CONTRIBUIÇÕES	35
1.6 ESTRUTURA DO TRABALHO	36
2. CONTROLE ESTATÍSTICO DE PROCESSOS E GRÁFICOS DE CONTROLE SHEWHART PARA ATRIBUTOS	37
2.1 CONTROLE ESTATÍSTICO DE PROCESSOS	37
2.2 VISÃO GERAL DOS GRÁFICOS DE CONTROLE	40
2.3 GRÁFICOS DE CONTROLE PARA ATRIBUTOS	42
2.4 GRÁFICOS DO TIPO SHEWHART PARA FRAÇÃO NÃO CONFORME (“P”) E NÚMERO DE NÃO CONFORMES (“NP”)	43
2.5 MEDIDAS DE DESEMPENHO	46
2.6 REGRAS SENSIBILIZANTES OU SUPLEMENTARES	48
2.7 AVANÇOS E MODIFICAÇÕES EM GRÁFICOS DO TIPO SHEWHART “P” OU “NP”	50
2.8 SUPOSIÇÕES	50
2.9 SÍNTESE DO CAPÍTULO	54
3. GRÁFICOS DE CONTROLE DE SOMA ACUMULADA	55
3.1 GRÁFICOS DE SOMA CUMULATIVA (CUSUM)	55
3.1.1 Avaliação de desempenho de gráficos CUSUM.....	58
3.2 CUSUM E SPRT	59
3.3 GRÁFICOS CUSUM PARA ATRIBUTOS	61
3.4 CUSUM PARA VARIÁVEIS COM DISTRIBUIÇÃO BINOMIAL ...	63
3.5 PARÂMETROS DE UM CUSUM BINOMIAL	64
3.6 VALOR DE REFERÊNCIA K DO CUSUM BINOMIAL	65
3.6.1 Outras considerações sobre k.....	68
3.7 O ARL E LIMITE SUPERIOR DE CONTROLE	69
3.7.1 Uso de Cadeias de Markov para calcular as medidas de desempenho	70
3.8 TÉCNICAS DE SIMULAÇÃO	73
3.8 APROXIMAÇÃO CD (“CORRECTED DIFFUSION”).....	74
3.9.1 A aproximação de Siegmund	74
3.9.2 Aproximação CD para distribuições de Bernoulli e Binomial	76
3.9.3 Fórmulas para calcular o ARL com a correção de h	78
3.9.4 Acurácia.....	80
3.10 CONSIDERAÇÕES FINAIS.....	81
3.11 SÍNTESE DO CAPÍTULO	82
4. GRÁFICOS COMBINADOS SHEWHART-CUSUM	83
4.1 GRÁFICOS COMBINADOS SHEWHART-CUSUM	83

4.2 GRÁFICOS COMBINADOS PARA ATRIBUTOS	88
4.3 CÁLCULO DOS PARÂMETROS DO GRÁFICO COMBINADO	89
4.3.1 ARL de um gráfico combinado Shewhart CUSUM.....	89
4.4 SÍNTESE DO CAPÍTULO.....	93
5. APROXIMAÇÕES PARA O LIMITE SUPERIOR DE UM CUSUM BINOMIAL	95
5.1 APROXIMAÇÃO DO LIMITE SUPERIOR DO CUSUM	95
5.2 ACURÁCIA DAS APROXIMAÇÕES	99
5.3 COMPARAÇÃO COM VALORES DA LITERATURA	103
5.4 COMPARAÇÃO COM VALORES SIMULADOS.....	105
5.5 USO DA APROXIMAÇÃO CD COM INTERPOLAÇÃO	109
5.6 CONSIDERAÇÕES FINAIS	116
5.7 SÍNTESE DO CAPÍTULO.....	116
6. DESEMPENHO DE UM GRÁFICO COMBINADO SHEWHART-CUSUM BINOMIAL	117
6.1 DISTRIBUIÇÃO DO RL DO GRÁFICO COMBINADO.....	117
6.1.1 RL do gráfico combinado - Exemplo 6.1	118
6.1.2 Resultados – Exemplo 6.1.....	119
6.1.3 RL do gráfico combinado - exemplo 6.2	130
6.1.4 Algumas considerações sobre o RL do combinado	133
6.2 DESEMPENHO DO GRÁFICO COMBINADO	133
6.2.1 Comparação de desempenho – Exemplo 6.3	134
6.3 COMPARAÇÃO DE DESEMPENHO – MAIS EXEMPLOS	138
6.4 SOBRE O DESEMPENHO.....	149
6.5 UMA TABELA COM PARÂMETROS PARA O GRÁFICO COMBINADO SHEWHART-CUSUM BINOMIAL	151
6.6 SÍNTESE DO CAPÍTULO.....	152
7. COMPARAÇÃO DO GRÁFICO COMBINADO COM PROCEDIMENTOS CUSUM E SHEWHART PARA SINALIZAR GRANDES ALTERAÇÕES.....	153
7.1 PROCEDIMENTOS METODOLÓGICOS.....	153
7.2 EXEMPLO 7.1	153
7.3 EXEMPLO 7.2	158
7.4 COMPARAÇÃO COM O USO DE RIR	161
7.5 ANÁLISE DOS RESULTADOS	163
7.6 SÍNTESE DO CAPÍTULO.....	164
8. PLANEJAMENTO DE UM GRÁFICO COMBINADO E APLICAÇÕES	165
8.1 PASSOS PARA CONTRUÇÃO DO GRÁFICO COMBINADO.....	165
8.1.1 Suposições.....	167
8.1.2 Outras distribuições	167
8.2 APLICAÇÃO 8.1: XÍCARAS DEFEITUOSAS	168
8.2 APLICAÇÃO 8.2: INDICADOR DE COLIFORMES TOTAIS.....	171
8.2.1 Verificação das Suposições.....	172

8.2.2 Construção e Aplicação do Gráfico Combinado	173
8.2.3 Comparação com outros gráficos	174
8.3 APLICAÇÃO 8.3: PROCESSO DE ROTULAGEM DE GARRAFAS	177
8.3.1 Verificação das suposições	178
8.3.2 Aplicação do Gráfico Combinado	179
8.3.3 Comparação com outros gráficos	180
8.4 SÍNTESE DO CAPÍTULO	183
9. CONCLUSÕES E CONSIDERAÇÕES FINAIS	185
9.1 CONTRIBUIÇÕES	188
9.2 DIFICULDADES ENCONTRADAS	189
9.3 SUGESTÕES PARA TRABALHOS FUTUROS	189
REFERÊNCIAS	193
APÊNDICE A – Rotina no R para aproximação do limite superior de um CUSUM binomial usando a função LambertW	213
APÊNDICE B – Rotina no R para aproximação do limite superior de um CUSUM binomial com a aproximação CD, resolvendo a Equação com um método iterativo	214
APÊNDICE C – Roteiro e Rotina no R para as simulações do Capítulo 5	215
APÊNDICE D – Rotina no R para aproximação do limite superior de um CUSUM binomial, com a função LambertW e interpolação logarítmica	217
APÊNDICE E – Roteiro e Rotina no R para as simulações do Capítulo 6	219
APÊNDICE F - Rotina no R para construção do gráfico Combinado Shewhart-CUSUM binomial unilateral superior	222
APÊNDICE G – Rotina no R para construção do gráfico CUSUM binomial unilateral superior	223
APÊNDICE H – Rotina no R para as aplicações do Capítulo 8..	224
APÊNDICE I – Verificação da aderência	231
ANEXO A – Dedução do valor de referência k do SPRT	235
ANEXO B – Valor de referência k para as distribuições Normal e de Poisson	237
ANEXO C – Aproximação de Siegmund	238
C.1 CORREÇÃO DO LIMITE DE DECISÃO	241
ANEXO D – Função findH do R	244
ANEXO E – Gráficos violin plots	246
ANEXO F – Conjunto de dados do exemplo de aplicação 8.2: Número de amostras não conformes	248
ANEXO G – Conjunto de dados do exemplo 8.3: Processo de rotulagem de garrafas	249
ANEXO H – Aderência à distribuição binomial	250

1. INTRODUÇÃO

A qualidade sempre esteve presente, como um desafio, em praticamente todos os produtos e serviços. É um dos mais importantes fatores de decisão dos consumidores na seleção de empresas e organizações. Mesmo onde não há concorrência direta é inegável a importância da qualidade. Portanto, uma organização que deseja gerar produtos e serviços com qualidade deve investir em atividades permanentes de monitoramento e controle.

As atividades de monitoramento e controle têm por objetivo principal a busca da qualidade e a melhoria contínua dos produtos e serviços. Como ação voltada à melhoria da qualidade cita-se a busca na eliminação sistemática de desperdícios diversos que, como consequência pode resultar em menores custos, maior produtividade, aumento na satisfação do consumidor, melhor reputação da organização, e também lucros maiores, quando for o caso.

Estas atividades de monitoramento e controle necessitam de técnicas apropriadas para que tenham sucesso. Além disso, a conscientização da necessidade do uso de métodos formais para este fim têm tido um desenvolvimento evolutivo. Como métodos formais, entendem-se métodos estatísticos e outras técnicas de resolução de problemas.

Os métodos estatísticos e sua aplicação na melhoria da qualidade têm uma longa história. Na década de 20, Walter Shewhart desenvolveu o conceito estatístico de gráficos de controle, dando um início formal ao campo do controle estatístico de qualidade. Os gráficos eram simples, voltados à atividade de inspeção, direcionado ao “chão de fábrica”.

Os gráficos de controle utilizados no monitoramento de características de qualidade de interesse representadas por variáveis contínuas, originárias de um processo de medição, são chamados de gráficos de controle para variáveis. Citam-se aqui gráficos para a média, a amplitude e para a variância de um processo. Em muitos processos, no entanto, as características de qualidade de interesse não podem ser mensuradas por grandezas físicas. Nestes casos, a qualidade pode ser representada ou pela proporção média de produtos defeituosos produzidos ou pela frequência de ocorrência de defeitos. Os gráficos de controle usados para monitorar estas medidas de qualidade são denominados gráficos de controle por atributos.

Os gráficos de Shewhart, para variáveis e atributos, alcançaram sucesso devido à sua simplicidade, em que a facilidade da regra de

decisão se baseia apenas no exame do último ponto analisado, ou seja, se ele está além dos limites de controle do gráfico deve-se intervir no processo. Esta é também uma grande desvantagem, pois ignora qualquer informação passada dada pela sequência anterior de pontos. Diz-se que o gráfico “não possui memória”. Isto torna o gráfico do tipo Shewhart relativamente insensível a pequenas mudanças no processo, da ordem de 1,5 desvios padrão ou menos (MONTGOMERY, 2004).

Dado este problema o gráfico de controle das Somas Acumuladas (CUSUM) e o gráfico de controle da Média Móvel Ponderada Exponencialmente (EWMA) foram desenvolvidos para corrigir essa lacuna. São indicados para o monitoramento de processos sujeitos a pequenas alterações. Nestes gráficos a decisão sobre o estado do controle estatístico do processo é baseada na informação acumulada das diversas amostras anteriores, e não somente na última delas, sendo possível sinalizar com maior rapidez os pequenos desajustes, assim como identificar no tempo o momento em que ocorre uma mudança no processo.

O desempenho de um gráfico de controle é medido pela rapidez com que este detecta alterações no processo. É comumente avaliado a partir de parâmetros relacionados com a distribuição do tempo necessário para o gráfico emitir um sinal. O número médio de amostras coletadas até emissão de um sinal (ARL-Average Run Length) é um desses parâmetros. Corresponde ao valor esperado (média) do RL (*run length* ou comprimento da corrida), número de observações que devem ser plotados antes de um ponto indicar uma condição fora de controle estatístico. Num gráfico de controle o ARL_0 indica o número médio de amostras coletadas até a emissão de um sinal durante o período sob controle. Já o ARL_1 representa o número médio de amostras coletadas até a emissão de um sinal que indique uma situação de fora de controle (ALVES e SAMOHYL, 2004).

Os gráficos do tipo Shewhart são considerados imbatíveis para a detecção de grandes desvios na média do processo, ou de aumentos significativos na variância ou da fração defeituosa (COSTA; EPPRECHT; CARPINETTI, 2004). No entanto, ainda de acordo com os autores, eles perdem rapidamente a eficiência na medida em que os processos vão ficando mais robustos. Outros critérios podem ser usados no gráfico do tipo Shewhart, como testes para sequências e o uso de limites de alerta. Todavia a utilização desses critérios, chamados de regras suplementares, reduz a simplicidade na interpretação do gráfico do tipo Shewhart.

Apesar dos gráficos CUSUM serem considerados pela literatura da área mais rápidos na detecção de mudanças de pequena e média magnitude que os gráficos Shewhart, estes últimos, ainda, são mais populares. Uma das razões é o fato do gráfico CUSUM não ser de fácil implementação e a caracterização do respectivo desempenho não ser necessariamente trivial. Calcular os limites convencionais de um gráfico Shewhart é simples, enquanto que no caso de um procedimento CUSUM são necessários métodos mais sofisticados para determinação dos parâmetros e medidas de desempenho.

Os gráficos do tipo Shewhart são recomendados na detecção de grandes alterações na média e variabilidade de um processo e gráficos CUSUM na sinalização de pequenas a moderadas alterações. Todavia se magnitude da mudança for desconhecida ou se modifica ao longo do tempo, nenhum dos gráficos mencionados irá ter um bom desempenho para todas as situações. Um procedimento possível é combinar múltiplos gráficos para abranger mudanças de diversas magnitudes, como por exemplo, adicionar limites de controle do tipo Shewhart aos gráficos CUSUM, para aumentar a sensibilidade deste para alterações maiores.

1.1 JUSTIFICATIVA

Gráficos combinados Shewhart-CUSUM têm como finalidade principal aumentar a sensibilidade dos gráficos CUSUM para mudanças maiores. Eles foram, na sua essência, planejados para possibilitar a detecção de mudanças de qualquer magnitude do parâmetro de interesse, sejam elas pequenas ou grandes. Gráficos combinados para monitoramento de médias são mais populares e já são aplicados no monitoramento ambiental.

Há, no entanto, pouquíssimos trabalhos abrangendo gráficos combinados Shewhart-CUSUM para dados com distribuição binomial. Existem apenas três referências bibliográficas, Morais (2002), Morais (2004) e Morais (2006), sendo que apenas uma delas (MORAIS, 2006) é um artigo em periódico, o que se configura uma grande lacuna nesta área. Várias questões permeiam seu desempenho e aplicabilidade. A dificuldade de implementação e caracterização do desempenho podem ser razões desta lacuna. As particularidades da distribuição binomial, principalmente a assimetria, reforçam estas deficiências. Neste sentido uma metodologia alternativa para a aproximação do limite superior de

controle de um gráfico CUSUM é proposta, motivada pela idéia de melhorar este passo.

Quanto aos procedimentos combinados, em relação ao monitoramento da média, a adição dos limites de controle de Shewhart aumenta efetivamente a sensibilidade de um gráfico CUSUM. Mas, ainda não se tem conclusões específicas sobre o que a adição das linhas Shewhart representaria para um CUSUM binomial. Morais e Pacheco (2006) analisam o desempenho de um gráfico combinado comparando-o com esquemas individuais, mas sem fixar uma taxa de alarmes falsos para todos os procedimentos, o que será feito neste trabalho. Assim, é possível avaliar, para as mesmas condições de planejamento, o que representa a adição de um limite superior de Shewhart a um CUSUM binomial. Para complementar, o gráfico combinado será comparado a alguns procedimentos planejados para detectar alterações maiores. Para o monitoramento de atributos não foi encontrada na literatura uma delimitação bem específica em termos de faixas específicas de magnitudes de mudança. Wu, Jiao e Lu (2008) consideram como uma grande alteração aquela com valores superiores um aumento de 100% na proporção sob controle, ou seja, $2p_0$ ($s = 100\%$), o que será adotado neste trabalho.

Quanto a aplicações a dados reais, não há registros, até a data de redação desta tese, da utilização de um gráfico combinado Shewhart-CUSUM binomial para este fim.

Este trabalho, portanto, é relevante para o estudo e avanço desta área. Pensando em estender o que foi desenvolvido neste trabalho àqueles que estiverem interessados, os procedimentos computacionais foram desenvolvidos no ambiente GNU R (R DEVELOPMENT CORE TEAM, 2009) e os códigos estão disponíveis nos Apêndices.

1.2 OBJETIVOS PROPOSTOS

Como objetivo geral busca-se o aperfeiçoamento e desenvolvimento do gráfico combinado Shewhart-CUSUM para dados com distribuição binomial. Para alcançar este propósito alguns objetivos específicos foram delineados e incluem:

a) Criar um modelo matemático para aproximar o limite superior de um gráfico CUSUM binomial, com base na aproximação de Siegmund e nas extensões do trabalho deste;

- b) avaliar a acurácia dos resultados obtidos com o modelo desenvolvido para aproximar o limite superior do CUSUM binomial;
- c) investigar o RL de um gráfico combinado mediante simulação e posterior análise exploratória de dados;
- d) analisar se, quando aplicado, o gráfico combinado atende aos parâmetros de projeto;
- e) verificar, a partir de simulação, se para dados com distribuição binomial a adição de linhas de Shewhart aumenta a sensibilidade de um gráfico CUSUM;
- f) avaliar o desempenho do gráfico proposto comparando-o com os gráficos Shewhart individuais e procedimentos CUSUM para alterações maiores;
- g) propor uma sequência de procedimentos para a construção de um gráfico combinado Shewhart-CUSUM binomial, examinando as suposições necessárias para o uso deste;
- h) elaborar funções no R para a aproximação proposta e para a construção de um gráfico combinado;
- i) aplicar o gráfico combinado a dados reais.

1.3 MÉTODO

Este trabalho é uma pesquisa aplicada, pois gera conhecimentos para aplicação prática e direcionados à solução de problemas específicos. A forma de abordagem é quantitativa, pois para alcançar os objetivos propostos são analisados os resultados obtidos. Quanto aos procedimentos técnicos, este trabalho é uma pesquisa bibliográfica, feita com material já publicado, como livros, artigos em periódicos, além de material disponibilizado na Internet. A pesquisa é também experimental, pois são desenvolvidos programas computacionais, realizadas simulações e aplicações.

Como recurso computacional foi utilizado o ambiente GNU R (R DEVELOPMENT CORE TEAM, 2009) na maior parte do trabalho, seja para o tratamento estatístico dos dados, como na elaboração das rotinas e programas. O software Mathematica foi aplicado para auxiliar nos procedimentos algébricos necessários (presentes no Capítulo 5). Para melhor entendimento do trabalho, em seguida, são delineados os procedimentos metodológicos utilizados neste trabalho para alcançar os objetivos propostos, relacionando com o capítulo correspondente.

a) Aproximação do limite superior de um gráfico CUSUM binomial (Capítulo 5) - o desenvolvimento da aproximação do limite superior h de um CUSUM binomial segue a metodologia apresentada por Rogerson (2006). Rogerson (2006) aplica a função LambertW e seu desenvolvimento em série na aproximação de Siegmund (1985) para calcular o ARL de um CUSUM para monitoramento da média de uma distribuição normal. Deste modo, Rogerson (2006) obteve uma fórmula direta para aproximação do limite superior h de um gráfico CUSUM. Neste trabalho a função de LambertW, e seu desenvolvimento em série, é aplicada à fórmula de Reynolds e Stoumbos (2001) para a aproximação do ARL_0 de um CUSUM binomial, com origem no trabalho de Siegmund (1985). Em seguida verifica-se a acurácia da aproximação, a partir da comparação dos resultados encontrados com os de trabalhos existentes e também com valores resultantes de processos de simulação. Foram elaboradas rotinas no R para a aproximação do limite superior h e para as simulações feitas.

b) RL do gráfico combinado Shewhart-CUSUM (Capítulo 6) - a análise do RL de um gráfico combinado Shewhart-CUSUM é feita a partir de simulação. As situações sob e fora de controle estatístico são simuladas e um gráfico combinado aplicado às mesmas. O procedimento é replicado r vezes. O RL, número da amostra onde ocorre a emissão do primeiro alarme para cada repetição, é registrado e armazenado. É feita a análise exploratória de dados, sendo calculadas algumas medidas descritivas, como a média (ARL), além de gráficos, como histogramas e o violin plot. Deste modo tem-se uma idéia de como se comporta o RL de um gráfico combinado. Foram elaboradas rotinas no R para as simulações.

c) Comparação do desempenho de um gráfico combinado (Capítulos 6 e 7) - o gráfico combinado é comparado com os gráficos individuais CUSUM e tipo Shewhart, também a partir de simulações. As situações sob e fora de controle estatístico são simuladas e os gráficos aplicados. O procedimento é replicado r vezes. O RL, número da amostra onde ocorre a emissão do primeiro alarme, é registrado e armazenado para cada repetição. É feita a análise exploratória de dados, sendo calculadas algumas medidas descritivas. Os valores de ARL, para diversas magnitudes de mudanças, são comparados graficamente. Esta comparação é feita analisando-se os gráficos dos valores de ARL construídos e também por meio do $rARL$, uma medida sugerida por Wu (2008) para este fim. Duas situações distintas são abordadas: magnitudes de mudança (de projeto) iguais para os gráficos CUSUM e combinado e, magnitudes maiores (de projeto) para o gráfico CUSUM.

d) Metodologia para a construção de um gráfico combinado (Capítulo 8) – é elaborada uma seqüência de procedimentos para a construção de um gráfico combinado Shewhart-CUSUM para dados com distribuição binomial. Através de pesquisa bibliográfica foram levantadas as suposições necessárias à aplicação do gráfico. Alternativas ao não atendimento das suposições são sugeridas. A metodologia proposta envolve a determinação dos parâmetros do gráfico combinado, sugerindo inclusive como calculá-los.

e) Aplicações (Capítulo 8) - o gráfico combinado desenvolvido é aplicado para dados da literatura existente e também em dados reais. Utiliza-se a metodologia de elaborada neste trabalho, com a aproximação do limite superior h do CUSUM binomial desenvolvida. Os resultados do gráfico combinado são comparados aos resultados de gráficos individuais tanto CUSUM quanto Shewhart. Foram elaborados e aplicados programas no R para a construção do gráfico combinado, do CUSUM binomial e do gráfico tipo Shewhart. Além disso, foram utilizadas funções do R para a verificação das suposições (aderência à distribuição binomial, autocorrelação e presença de superdispersão).

1.4 LIMITAÇÕES

Neste trabalho são propostas algumas contribuições ao estudo dos gráficos de controle combinados Shewhart-CUSUM para dados com distribuição binomial. A metodologia desenvolvida, devido à complexidade, está restrita apenas aos gráficos unilaterais superiores.

1.5 CONTRIBUIÇÕES

Este trabalho visa contribuir para o avanço das técnicas de Controle de Estatístico de Processo. Considerando-se os objetivos propostos neste trabalho, tem-se, que as principais contribuições residem nos seguintes pontos:

a) Desenvolvimento de uma nova abordagem para a aproximação do limite superior de um CUSUM binomial;

b) determinação, mediante simulação, de valores de ARL para os gráficos combinados Shewhart-CUSUM binomiais. É um primeiro

trabalho com valores que podem ser utilizados para futuras comparações;

c) elaboração de uma metodologia para construção de um gráfico combinado Shewhart-CUSUM binomial, envolvendo verificação das suposições das suposições necessárias, e sugestão de alternativas para o caso de estas não serem atendidas;

d) elaboração de rotinas no R para a aproximação do limite superior, gráfico CUSUM binomial e gráfico combinado Shewhart-CUSUM binomial;

e) aplicações do gráfico Shewhart-CUSUM binomial a dados reais, onde os conhecimentos teóricos são transferidos a situações práticas.

1.6 ESTRUTURA DO TRABALHO

O trabalho está dividido em nove Capítulos. O Capítulo 2 traz uma revisão conceitual sobre Controle Estatístico de Qualidade, os gráficos Shewhart e as suposições para aplicação deste. No Capítulo 3 é realizada uma revisão do gráfico CUSUM. A ligação do CUSUM com o SPRT, o CUSUM para distribuições pertencentes à família exponencial, e suas propriedades com relação ao valor de referência k . O limite de decisão h e medidas de desempenho também são exploradas, assim como as suas especificidades. No Capítulo 4 é feita uma revisão bibliográfica dos gráficos combinados Shewhart-CUSUM. O Capítulo 5 apresenta a proposta de formulação para aproximar o limite superior h de um CUSUM binomial. O Capítulo 6 traz algumas análises da distribuição do RL do gráfico combinado e avalia o desempenho deste sob a ótica de aumento de sensibilidade do CUSUM. No Capítulo 7 o gráfico combinado é comparado, em termos de performance com gráficos CUSUM planejados para detectar alterações maiores. No Capítulo 8 são apresentadas três aplicações, uma a dados adaptados da literatura e duas situações reais: monitoramento de um indicador de qualidade de água potável e um processo produtivo de rotulagem de garrafas. Finalizando, no Capítulo 9, estão as conclusões e sugestões para trabalhos futuros.

2. CONTROLE ESTATÍSTICO DE PROCESSOS E GRÁFICOS DE CONTROLE SHEWHART PARA ATRIBUTOS

Neste capítulo será apresentada uma visão geral dos conceitos relacionados ao controle estatístico de processos e principais gráficos de controle para atributos. Na Seção 2.1 é feita uma revisão teórica sobre Qualidade e Controle Estatístico de Processos e na Seção 2.2 está uma visão geral dos Gráficos de Controle. A Seção 2.3 aborda os Gráficos de Controle para atributos, abordando os tradicionais gráficos de Shewhart “p” e “np”. Os parâmetros de desempenho deste procedimento são descritos na Seção 2.4. A Seção 2.5 trata dos limites probabilísticos e a seção 2.6 das regras suplementares (ou sensibilizantes). Na Seção 2.7, são listados alguns trabalhos recentes que envolvem modificações dos gráficos Shewhart “p” e “np”. Na Seção 2.8 estão as suposições necessárias para a aplicação destes gráficos.

2.1 CONTROLE ESTATÍSTICO DE PROCESSOS

Qualidade é um termo constante nas organizações. Fala-se em produtos e serviços com qualidade, gestão da qualidade, qualidade total, entre tantos outros termos. Não existe, no entanto, na literatura uma definição única para qualidade (COSTA; EPPRECHT; CARPINETTI, 2004). Juran e Gryna (1992) conceituou qualidade como “adequação ao uso”. Qualidade, conforme Paladini (2000), é uma relação da organização com o mercado, e, dentro deste contexto, pode ser definida como uma relação de consumo. O conceito de qualidade é, por assim dizer, dinâmico. Deste modo, o conceito de qualidade ainda pode ser definido como uma relação do mercado e a sociedade. Portanto “produzir” ou “oferecer” qualidade implicam atividades permanentes de monitoramento e controle.

O controle permanente dos processos é uma condição básica para a manutenção da qualidade de bens e serviços. Entenda-se por processos, os industriais que geram bens de consumo e os de organizações que prestam serviços. Nesta segunda classificação estão bancos, prefeituras, companhias de águas e esgoto, hospitais e clínicas, por exemplo. O monitoramento de processos, se realizado de forma inteligente, implica em custos que podem ser pagos facilmente, pois qualidade agrega valor (COSTA; EPPRECHT; CARPINETTI, 2004).

Particularmente, para a engenharia, onde a mensuração tem um papel vital, o monitoramento visa verificar a contínua estabilidade dos processos. Medidas necessitam ser acuradas, consistentes e os valores próximos aos reais. Há a necessidade de se detectar alterações no processo, de modo que os operadores possam iniciar a busca do problema e tomar ações corretivas (BERTHOUEX e BROWN, 2002)¹.

O Controle Estatístico de Processo (CEP) é uma técnica formal eficiente e poderosa utilizada para o monitoramento de processos. Compõe-se de uma coleção de ferramentas de resolução de problemas útil na obtenção da estabilidade de um processo. O Gráfico de Controle é uma dessas ferramentas e, provavelmente a mais sofisticada tecnicamente (MONTGOMERY, 2004).

O início formal do Controle Estatístico de Processos deu-se em meados de 1924, quando Walter A. Shewhart desenvolveu e aplicou os gráficos de controle na empresa *Bell Telephone Laboratories*. Era uma ferramenta simples para uso no “chão de fábrica”, por operários com pouca instrução. Para ele, todo e qualquer processo, por melhor que seja projetado e controlado, possui uma variabilidade natural inerente ao processo, proveniente de causas aleatórias. Quando um processo apresenta apenas esta variabilidade natural, diz-se que está no estado de controle estatístico, ou sob controle. Além das causas aleatórias, os processos podem sofrer a ação de causas especiais e estruturais, resultando na ocorrência de uma variabilidade extra. Essas causas aumentam a dispersão e podem até afastar a característica de qualidade de interesses do seu valor em controle (MONTGOMERY, 2004; COSTA; EPPRECHT; CARPINETTI, 2004).

Uma causa especial é assinalável, imprevisível e, em geral, única. No entanto é suficientemente grande para produzir perturbações fortes em um processo. Exemplos de causas especiais são: substância estranha na matéria prima, treinamento inadequado, para o qual faltou um ensinamento importante, entre outros. Essas causas devem ser identificadas e eliminadas ou reduzidas por ações compensatórias (SAMOHYL, 2009).

Quando, além das causas aleatórias, essas causas especiais estão presentes, diz-se que o processo está fora de controle estatístico. Em relação a tal aspecto, entende-se que os processos devem ser monitorados para detectar a ocorrência destas causas especiais e permitir, no menor intervalo de tempo possível, o emprego de ações corretivas (COSTA; EPPRECHT; CARPINETTI, 2004).

¹Todas as citações e referências em língua inglesa neste trabalho são traduções do autor.

Outra causa é a chamada estrutural. Esta também é eliminável ou compensável. Ocorre periodicamente ou em função de outra variável relacionada com a causa identificada. A relação entre variável e causa é previsível e de natureza repetitiva. Quando o período entre as ocorrências for grande, esta causa pode se confundir com uma causa especial. Como exemplo, a produtividade dos funcionários pode ser afetada nos dias subseqüentes a importantes jogos de futebol (SAMOHYL, 2009).

A má qualidade em um processo é resultado de causas especiais, estruturais e também das causas aleatórias (ou comuns). As causas comuns são relativamente pequenas, mas ocorrem em grande número, afetando a variabilidade do processo. Os gráficos de controle não são, de modo geral, apropriados para a análise de causas comuns (SAMOHYL, 2009).

As causas comuns (ou aleatórias) podem ser reduzidas, mas não totalmente eliminadas. Irão sempre existir, em virtude da variabilidade natural inerente ao processo. A diminuição da irregularidade das causas comuns pode ser alcançada, por exemplo, com investimentos em novas e melhores máquinas, matéria-prima melhor e treinamento intensivo de operadores, entre outros (SAMOHYL, 2009).

Assim, um Gráfico de Controle é uma ferramenta do CEP para identificar e analisar causas especiais e estruturais, com objetivo de posterior eliminação (preferencialmente) ou redução por meio de medidas compensatórias. O gráfico de controle é um elemento visual para o monitoramento de conformidade de características de produtos e processo (SAMOHYL, 2009).

Até pouco tempo, os gráficos de controle eram direcionados exclusivamente para o monitoramento de processos industriais. Da mesma forma foram ampliadas as aplicações dos gráficos de controle. O paradigma tradicional é o processo industrial, mas esta ferramenta já se estendeu para processos administrativos e de serviços (SAMOHYL, 2009). Algumas aplicações de gráficos de controle em outras áreas podem ser vistas em:

- a) Administração: Hawkins e Olwell (1998) – erros em documentos;
- b) Análises clínicas: Westgard *et al.*, (1977);
- c) Bioterrorismo: Frisen (2007), Meyer *et al.*, (2007);
- d) Procedimentos cirúrgicos: Naik; de Vito e Halpern (2003), Beiles e Morton (2004), Coulquhuon (2008) – avaliação de desempenho;

e) Educação: Savic (2006), Henning *et al.*,(2009) – avaliação de projetos;

f) Saúde Pública e Epidemiologia: Blacksell *et al.*,(1994), Hawkins e Olwell (1998), Sanches (2000), Morton *et al.*, (2001) Sonesson e Bock (2003), Woodall (2006);

g) Meio ambiente: Gibbons (1999), Barrat *et al.*, (2007), Mesnil e Petitgas (2009) – monitoramento de aquíferos em aterros sanitários, emissões atmosféricas e indicadores em ecossistemas marinhos respectivamente.

Outras aplicações em diversas áreas, ainda, podem ser citadas, como gestão de pessoas, análise e prevenção de crimes, detecção de fraudes, transações financeiras, ciências do ambiente e biologia (STOUMBOS *et al.*, 2000; MORAIS, 2002; MORAIS e PACHECO, 2006).

2.2 VISÃO GERAL DOS GRÁFICOS DE CONTROLE

Um gráfico de controle é uma ferramenta estatística visual que desperta para a presença de causas especiais (SAMOHYL, 2009). Consiste basicamente na plotagem de linhas que representam os limites de controle superior (LSC) e inferior (LIC), a média ou alvo do processo (linha central - LC) e dos pontos observados, que representam, em geral, uma estatística relacionada à variável de interesse. Se um ou mais pontos estiverem além dos limites de controle há indicação de que o processo possa estar fora de controle estatístico, isto é, pode existir um problema. Um gráfico de controle em formato conceitual pode ser visto na Figura 2.1.

Esta variável, denominada também de característico da qualidade, pode ser de natureza física, sensorial ou temporal. Para avaliação dela é feita uma classificação ou uma medição. Diz-se que o característico da qualidade é uma variável se resulta de uma medição e, é um atributo, se resulta de uma classificação (VIEIRA, 1999). Assim, os gráficos de controle podem ser classificados em gráficos de controle para variáveis e gráficos de controle para atributos.

Para o desenvolvimento de um gráfico de controle são necessárias duas fases. A primeira fase, denominada de fase I, é em geral um estudo retrospectivo. Consiste em obter-se uma amostra

representativa dos dados com o objetivo de determinar os limites de controle.

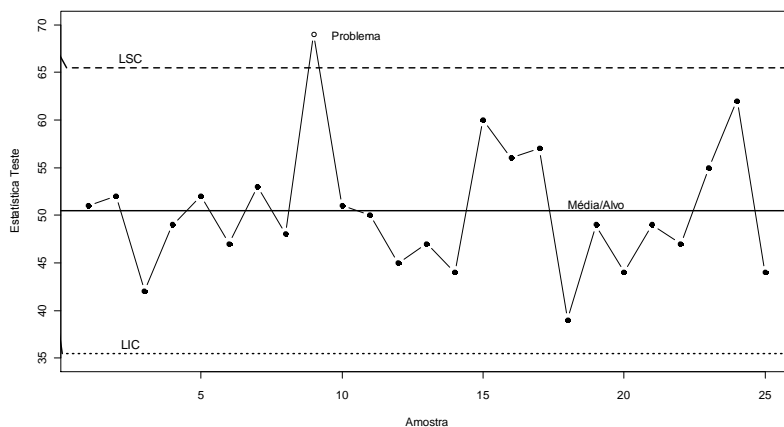


Figura 2. 1 – Gráfico de Controle do tipo Shewhart.

Fonte: Adaptado de Samohyl (2009).

Esta fase, que corresponde à estimação de parâmetros do processo, só deve ser encerrada quando se tem a certeza de que o processo encontra-se estável e ajustado. A segunda, chamada de fase II, tem o intuito de monitoramento do processo e utiliza os limites especificados na etapa anterior (MASON e YOUNG, 2002; COSTA; EPPRECHT; CARPINETTI, 2004).

Na fase II pode surgir a dúvida sobre a necessidade e frequência de atualização dos limites de controle, para incorporar novos dados e rever o desempenho do procedimento. A resposta é subjetiva e depende de vários fatores: tempo, recursos e prioridades (SAMOHYL, 2009). Cada caso é único, com suas próprias particularidades.

A eficácia de um gráfico de controle é medida pela rapidez com que este dispositivo detecta alterações no processo. A análise da relação entre o custo de operação e a eficácia do gráfico de controle deve nortear a escolha de seus parâmetros: tamanho das amostras, a frequência da amostragem e o fator que estabelece o posicionamento dos limites de controle no gráfico (COSTA; EPPRECHT; CARPINETTI, 2004; MONTGOMERY, 2004).

Nas próximas seções, os gráficos para controle por atributos, incluindo os conceitos estatísticos em que estão norteados, serão descritos com mais detalhes.

2.3 GRÁFICOS DE CONTROLE PARA ATRIBUTOS

Gráficos de controle para atributos aplicam-se ao casos em que o característico da qualidade é representado por uma variável discreta, resultado de uma classificação e/ou contagem.

A qualidade de um produto (ou serviço) pode ser avaliada pela classificação em itens “defeituosos” (não conformes) ou “não defeituosos” (conformes). Desse modo, “ser defeituoso” ou “não defeituoso” é um atributo do processo. Os gráficos mencionados recebem a designação de “Gráficos de Controle por Atributos” (COSTA; EPPRECHT; CARPINETTI, 2004).

Gráficos por atributos podem ser aplicados para monitorar, por exemplo:

a) Número de peças defeituosas em uma linha de produção. Classificação: ser defeituoso ou não.

b) Número de clientes satisfeitos com um determinado serviço. Classificação: satisfeito ou não.

c) Número de infectados por doença específica. Classificação: estar ou não infectado.

d) Número de amostras de um efluente com presença de coliformes termotolerantes. Classificação: apresentar coliformes termotolerantes ou não.

e) Número de cirurgias necessárias para alcançar proficiência em determinada técnica cirúrgica. Classificação: realizar a cirurgia sem auxílio do tutor.

Controlar variáveis mensuráveis é mais eficaz que controlar atributos. A medição provê uma informação mais precisa com tamanhos menores de amostras, dando uma qualidade científica à atividade de monitoramento (WOODALL, 1997; VIEIRA, 1999; PALADINI, 2000; SAMOHYL, 2009)

Na prática, pode-se considerar que o controle por atributos possui algumas vantagens principalmente do ponto de vista econômico. A avaliação por classificação é mais rápida. Equipamentos e análises são tecnicamente mais simples, permitindo, em curto prazo um investimento menor.

Existem, portanto, diversas razões para evitar mensurações de algumas características do produto ou processo, concentrando-se as ações apenas na contagem de itens não conformes ou no número de

defeitos em uma peça. Muitas vezes o monitoramento de atributos é a primeira atividade assumida por uma empresa que inicia um programa de qualidade (SAMOHYL, 2009).

Para algumas atividades a categorização é inerente. Neste caso, as características da qualidade não podem ser medidas em escala numérica ou mesmo de forma quantitativa (WOODALL, 1997; WU; JIAO; LIU, 2008). São casos, por exemplo, de vigilância epidemiológica.

Gráficos para atributos são atrativos, tanto para processos industriais quanto para setores de serviços, pela sua simplicidade. Abrangem, de certo modo, o tratamento de múltiplas características da qualidade, com facilidade de entendimento por pessoas de diferentes níveis nas organizações, sobretudo pela prevalência de dados discretos em muitos setores (WU; JIAO; LIU, 2008; DURAN e ALBIN, 2009).

2.4 GRÁFICOS DO TIPO SHEWHART PARA FRAÇÃO NÃO CONFORME (“P”) E NÚMERO DE NÃO CONFORMES (“NP”)

Um gráfico de controle para atributos extensivamente utilizado é o da fração não conforme “p” (DURAN e ALBIN, 2009; SAMOHYL, 2009). Este gráfico serve, por exemplo, para monitorar processos que produzem certa porcentagem de produtos defeituosos (COSTA; EPPRECHT; CARPINETTI, 2004), sendo muito popular nas fábricas onde a utilização de controle estatístico de processo é, ainda, muito embrionária (SAMOHYL, 2009).

A fração não conforme é definida como a razão entre o número de itens não conformes em uma população e o total de itens dessa população. Os dados são classificados como tendo ou não o atributo – ser conforme/ não defeituoso. Esta classificação dos dados (ter ou não um determinado atributo) leva-os a aproximação pela distribuição binomial (HAWKINS e OLWELL, 1998).

Um gráfico de controle para a fração não conforme está associado à distribuição binomial. Nesta, a probabilidade de que um item qualquer seja classificado como defeituoso ou não conforme é p , constante, e os itens produzidos são variáveis aleatórias identicamente distribuídas. Cada unidade produzida é uma realização de uma variável aleatória de Bernoulli com parâmetro p (MONTGOMERY, 2004).

Se uma amostra aleatória de tamanho n 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 , ou seja,

$$P(D = x) = \binom{n}{x} p^x (1 - p)^{n-x}, x = 0, 1, \dots, n. \quad (2.1)$$

A fração amostral não conforme \bar{p}_s é definida como a razão entre o número de unidades não conformes na amostra D e o tamanho da amostra, isto é,

$$p_s = \frac{D}{n}. \quad (2.2)$$

A média do processo é \bar{p}_s , e o desvio padrão σ é calculado conforme a Equação 2.3.

$$\sigma = \sqrt{\frac{p_s(1-p_s)}{n}}. \quad (2.3)$$

Inicialmente, aborda-se o gráfico para a fração não conforme com limites de controle de Shewhart. Este gráfico é comumente conhecido como Gráfico de Shewhart do tipo “p”. Parte-se da forma geral do gráfico de controle de Shewhart (MONTGOMERY, 2004), ou seja,

$$\text{Limite Superior de Controle} \quad LSC = \mu_w + q_w \sigma_w; \quad (2.4)$$

$$\text{Linha Central} \quad LC = \mu_w \quad e \quad (2.5)$$

$$\text{Limite Inferior de Controle} \quad LIC = \mu_w - q_w \sigma_w, \quad (2.6)$$

onde w é a estatística que mede uma característica da qualidade; μ_w é média de w e σ_w é o desvio padrão de w .

Nas Equações 2.4, 2.5 e 2.6, q_w corresponde a distância dos limites de controle à linha central, em valores múltiplos de desvios padrão de w . Estes limites dependem das probabilidades de erro tipo I (α) e erro tipo II (β). A probabilidade α (erro tipo I) corresponde a considerar o processo fora de controle quando na verdade este está sob controle estatístico (alarme falso). Já a probabilidade de ocorrer β (erro tipo II) denota a probabilidade de se assumir que o processo está sob controle, sendo que não está (não detecção). Os limites de controle 3σ (três desvios padrão, $q_w = 3$) são, de modo geral, utilizados no gráfico de controle do tipo Shewhart para a fração não conforme.

Então, supondo que a verdadeira fração não conforme do processo, p_s , seja conhecida ou um valor especificado, a linha central e

os limites de controle (3σ) de um gráfico de controle para a fração não conforme p podem ser definidos como

$$\text{Limite Superior de Controle} \quad LSC = p + 3\sqrt{\frac{p(1-p)}{n}}; \quad (2.7)$$

$$\text{Linha Central} \quad LC = p \quad (2.8)$$

e

$$\text{Limite Inferior de Controle} \quad LIC = p - 3\sqrt{\frac{p(1-p)}{n}}. \quad (2.9)$$

Estes limites são aproximações da distribuição normal à binomial e algumas regras são sugeridas na literatura para aplicação deste tipo de gráfico. Uma regra, chamada aqui de regra 1, é $np > 5$ e outra, regra 2, é $np(1-p) > 9$ (SCHADER e SCHMID, 1989).

Se a fração não conforme do processo, p , não é conhecida pode ser estimada a partir dos dados observados. Selecionam-se m amostras preliminares de tamanho n . Deste modo se existem D_i unidades não conformes na amostra i , a fração não conforme na i -ésima amostra é dada como

$$\hat{p}_i = \frac{D_i}{n}, \quad i = 1, 2, \dots, m, \quad (2.10)$$

e a média dessas frações não conformes é:

$$\bar{p} = \frac{\sum_{i=1}^m D_i}{mn} = \frac{\sum_{i=1}^m \hat{p}_i}{m}. \quad (2.11)$$

A estatística \bar{p} estima a fração não conforme desconhecida, p . Sendo assim a linha central e os limites de controle do gráfico são os mesmos que nas Equações 2.7, 2.8 e 2.9, mas substituindo p por \bar{p} .

Um gráfico de controle pode-se basear no número de não conformes ao invés da fração não conforme. É uma forma de simplificar a utilização do gráfico “p” para tornar mais fácil, por exemplo, as tarefas de um operador de linha de produção (SAMOHYL, 2009). Este gráfico, chamado na literatura de “np”, é considerado mais fácil de interpretar que o gráfico “p” pelo pessoal de chão de fábrica (MONTGOMERY, 2004).

Os Limites de Controle (Superior e Inferior) e a Linha Central são dados por

$$LSC = np + 3\sqrt{np(1-p)}; \quad (2.12)$$

$$LC = np \quad (2.13)$$

e

$$LIC = np - 3\sqrt{np(1-p)}. \quad (2.14)$$

Do mesmo modo que no gráfico anterior, se não se dispõe de um valor padrão para p , então \bar{p} pode ser utilizado para estimar p .

Foi constatado matematicamente que os gráficos “p” e “np” são operacionalmente equivalentes, produzindo os mesmos sinais e alarmes falsos (SAMOHYL, 2009).

2.5 MEDIDAS DE DESEMPENHO

Uma das maneiras de se avaliar o desempenho de um gráfico de controle é mediante o ARL (*average run length* ou comprimento médio da corrida). O ARL corresponde ao valor esperado (média) do RL (*run length* ou comprimento da corrida), número de observações que devem ser plotados antes de um ponto indicar uma condição fora de controle estatístico. O RL é uma variável aleatória, tendo uma distribuição de probabilidades e momentos, como a média e a variância. Assim, a distribuição de probabilidades do RL de um gráfico do tipo Shewhart é uma distribuição geométrica com parâmetro $p = \alpha$. (MONTGOMERY, 2004; COSTA, EPPRECHT; CARPINETTI, 2004).

O ARL_0 corresponde ao número médio de amostras que são esperadas até a emissão de um alarme falso. O ARL_1 é o número médio de amostras esperadas até a sinalização de uma mudança. O alarme falso, como foi visto, está associado ao erro tipo I. A hipótese nula (H_0) neste caso, é que o processo é estável. A hipótese alternativa (H_1) implica em um processo não estável. Um gráfico de controle pode ser visto como uma seqüência de testes de hipóteses. Alarmes falsos são denominados ARL_0 em analogia com H_0 e verdadeiros de ARL_1 (H_1) (SAMOHYL, 2009). Pode-se determinar o ARL, em termos de “alarme

falso” (erro tipo I, probabilidade α), para o gráfico Shewhart de acordo com a Equação (2.15).

$$ARL_0 = \frac{1}{\alpha_s}. \quad (2.15)$$

Denota-se aqui por α_s a probabilidade em que cada ponto exceda os limites de controle, dado que o processo esteja sob controle estatístico.

EXEMPLO 2. 1: Para um gráfico de controle 3σ , o ARL_0 pode ser calculado como

$$ARL_0 = \frac{1}{\alpha_s} = \frac{1}{0,0027} \cong 370,$$

com $\alpha_s = 2(0,00135) = 0,0027$, conforme Figura (2.2). Pode-se observar que, se o processo está sob controle, tem-se, em média, um falso alarme a cada 370 amostras aproximadamente.

Se a frequência de alarmes falsos for considerada inaceitável, a alternativa consiste em alargar os limites de controle. Isto corresponde a passar q_w (Equações 2.12, 2.13 e 2.14), por exemplo, de 3,0 para 3,1. Assim, o risco de alarme falso diminui (COSTA; EPPRECHT; CARPINETTI, 2004). O risco α é então uma função apenas do fator de abertura dos limites de controle.

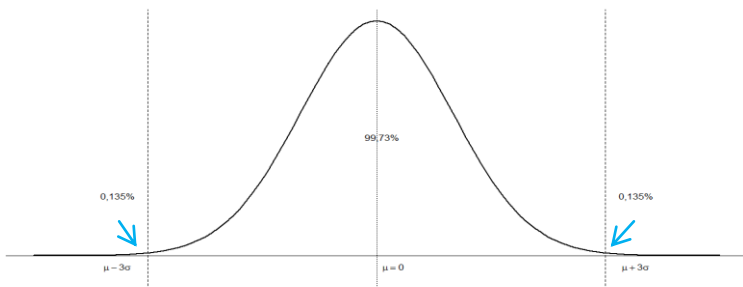


Figura 2. 2 – Probabilidade de ocorrência de observações além de três desvios padrão da média em uma distribuição normal.

Fonte: Adaptado de Coelho (2008).

O uso do ARL como medida para descrever o desempenho dos gráficos de controle não é consenso, tendo recebido críticas. Algumas razões apontam para o fato de que a distribuição de probabilidades do RL para um gráfico de controle Shewhart seja modelada justamente por uma distribuição geométrica. A distribuição geométrica é muito assimétrica, de modo que o valor esperado da distribuição não seja necessariamente um valor “típico” do ARL (MONTGOMERY, 2004).

Os gráficos “p” e “np” podem não alcançar a taxa de alarmes falsos desejada (ver Tabela 2.1), que equivale a não atingir o ARL_0 (DURAN e ALBIN, 2009).

Tabela 2. 1 –Valores calculados de ARL_0 para $n = 200, 600, 1000$ e $p_0 = 0,1$ e $0,001$.

n	p_0	ARL_0
200	0,1	294
600	0,1	441
1000	0,001	300

FONTE: adaptado de Albin e Duran (2009).

Pode ser verificado que para os valores da Tabela 2.1, tanto a regra 1 ($np > 5$) como a regra 2 ($np(1 - p)$) são atendidas; nos três casos o valor do ARL_0 difere significativamente do desejado.

Existem algumas opções para contornar este problema. Para os gráficos do tipo Shewhart, uma alternativa é usar os limites probabilísticos (ou exatos). Os limites são calculados com a distribuição binomial e não aproximando à normal. Estes limites modificam o gráfico de controle, considerando a real probabilidade que cada ponto tem de ultrapassá-los. Assim, são indicados para distribuições não simétricas, como a de Poisson e também muitos casos da binomial (WETHERILL e BROWN, 1991; XIE e GOH, 1993; SHORE, 2000; SHARMA, 2003; MONTGOMERY, 2004).

2.6 REGRAS SENSIBILIZANTES OU SUPLEMENTARES

Os gráficos de controle “np” de Shewhart são amplamente utilizados devido à sua simplicidade. No entanto, estes gráficos não têm

memória, sendo pouco efetivos para a detecção de mudanças mais moderadas.

As regras suplementares (ou sensibilizantes) aplicadas aos gráficos do tipo Shewhart procuram suplantar esta dificuldade, todavia reduzem a simplicidade e facilidade de uso (ROCHA, 2004). Essas regras foram criadas com o propósito de acelerar a detecção de alterações no processo. Muitas delas tentam formalizar a percepção de que algo, no processo, possa estar fora de ordem. A incorporação dessas regras implica no aumento do risco α , ou seja, há um aumento no número de alarmes falsos (COSTA; EPPRECHT; CARPINETTI, 2004).

Algumas regras sensibilizantes para os gráficos de controle de Shewhart são listadas a seguir. As quatro primeiras são conhecidas como regras da Western Electric, pois foram sugeridas no Western Electric Handbook (1956). A primeira regra listada corresponde ao sinal de ação padrão, não sendo uma regra sensibilizante.

Um processo é considerado fora de controle estatístico se um ou mais dos critérios abaixo forem verificados:

- 1) Um ou mais pontos fora do limite de controle;
- 2) dois ou três pontos sucessivos fora dos limites de alerta dois σ ;
- 3) quatro ou cinco pontos consecutivos além dos limites de um σ ;
- 4) uma sequência de oito pontos consecutivos de um mesmo lado da linha central;
- 5) seis pontos em uma sequência sempre crescente ou decrescente;
- 6) quinze pontos em sequência na zona C (um σ), tanto acima como abaixo da linha central;
- 7) quatorze pontos em uma sequência alternadamente para cima e para baixo;
- 8) oito pontos em uma sequência de ambos os lados da linha central com nenhum na zona C (um σ);
- 9) um padrão não usual ou não aleatório dos dados;
- 10) um ou mais pontos perto de um limite de alerta ou de controle.

Nelson (1994) e Acosta-Mejia (1999) apontam duas vantagens no uso de regras suplementares. Elas são capazes de sinalizar a ocorrência de uma causa especial indicando que a fração de defeituosos efetivamente diminuiu e a substituição do limite inferior de controle por regras suplementares pode resultar em um aumento geral no desempenho do gráfico Shewhart para valores pequenos de p .

2.7 AVANÇOS E MODIFICAÇÕES EM GRÁFICOS DO TIPO SHEWHART “P” OU “NP”

Diversos trabalhos publicados focam alterações no gráfico do tipo Shewhart com objetivos de contornar algumas limitações e aumentar sua eficiência, sem perder a simplicidade de uso e interpretação. Entre eles:

a) Gráficos de controle com tamanho de amostra variável para atributos (EPPRECHT e COSTA, 2001);

b) gráficos do tipo SHEWHART “np” baseados em valores específicos para a fração não conforme p_0 sob controle, decréscimo na proporção p_- , acréscimo na proporção p_+ , seus correspondentes desenhos de RL e probabilidades correspondentes (WU e LUO, 2003);

c) gráfico “np” ótimo com restrições, mantendo a taxa de alarmes falsos em um determinado nível (WU; LUO; ZHANG, 2006);

d) gráfico do tipo Shewhart modificado para detecção de melhoras em processos com proporção de itens não conformes muito pequena e o limite inferior calculado é negativo ou zero (LUCAS *et al.*, 2006). Estes gráficos modificados não serão abordados neste trabalho.

2.8 SUPOSIÇÕES

A aplicação dos gráficos de controle de Shewhart e CUSUM para dados com distribuição binomial só será eficaz se os dados atenderem às seguintes premissas: os dados devem ter distribuição binomial; não apresentar superdispersão e devem ser independentes.

O distanciamento da distribuição binomial pode afetar o desempenho dos gráficos de controle, tanto de Shewhart, como o CUSUM (ver Capítulo 3), resultando em mais alarmes falsos. Para Jones e Govindaraju (2000) é possível estabelecer limites Shewhart sem verificar a aderência, mas recomendam a verificação para o caso de aparecerem muitos pontos além dos limites sem nenhuma explicação pelo processo (por exemplo, alarmes falsos). Wetherill e Brown (1991), todavia, recomendam verificar a aderência antes de estabelecer os limites de controle.

A verificação da aderência pode ser feita por testes formais, como por exemplo, o teste qui quadrado (WETHERILL e BROWN, 1991; JONES e GOVINDARAJU, 2000; DEVORE, 2006). Uma breve

discussão sobre testes e de aderência à distribuição binomial está no Anexo I.

Já, a presença de superdispersão (variância maior que $np(1 - p)$) resulta também no aumento do número de alarmes falsos. Recomenda-se um teste estatístico formal para checar a superdispersão.

Procedimentos para verificar graficamente essas suposições (superdispersão e aderência) são propostos por Jones e Govindaraju (2000). Transformações nos dados, como a arco seno, podem ser usadas e, mediante um gráfico de probabilidade normal verifica-se a presença de superdispersão. Para a distribuição binomial, segundo os autores, se houver superdispersão a suposição de aderência à distribuição binomial não é justificada. A transformação Box-Cox não é considerada muito efetiva no caso de atributos (JONES e GOVINDAJARY, 2000).

A suposição mais importante relativa aos gráficos de controle é a independência das observações de um processo produtivo, ou seja, que não haja correlação entre as unidades amostrais. Se a suposição de normalidade for violada num grau moderado, os gráficos de controle (para variáveis mensuráveis) ainda funcionam razoavelmente, o que não acontece com a suposição de independência. Em relação às variáveis binomiais não se tem registros de artigos que tenham investigado com maior profundidade se a violação da aderência em um grau menor afeta o desempenho de um gráfico do tipo Shewhart “p” ou “np”. Mas, gráficos de controle convencionais não funcionam bem se a característica da qualidade apresenta níveis, ainda que pequenos, de correlação ao longo do tempo. Especificamente, com o processo sob controle, resulta em mais alarmes falsos (SHARMA, 2003).

Infelizmente, em muitos processos, principalmente industriais, essa suposição de independência nem sempre é satisfeita, mesmo aproximadamente. Isto é comum na maioria dos casos de processos automatizados, principalmente quando o tamanho da amostra é um (1) e o intervalo entre amostras é muito pequeno. Assim as observações podem ser autocorrelacionadas (SHARMA, 2003). Alwan e Roberts (1995) mostram que a correlação entre unidades amostrais aparece com uma frequência bem maior do que se pensa. Na realidade, a correlação entre as unidades amostrais, no que se refere à variável resposta em estudo, tende a 1(um) quando o intervalo entre inspeções tende a zero. Deste modo, uma das alternativas sugeridas para tratamento da correlação é um maior espaçamento entre as unidades amostrais (ou grupos racionais) selecionadas para inspeção. No entanto, esta alternativa pode ocasionar grandes perdas financeiras uma vez que pelo

fato de se espaçar demais as inspeções demora-se mais para se detectar um problema no processo (MINGOTI e CARVALHO, 2003).

Amostras ou unidades coletadas em sequência consistem numa série temporal de contagens. Um primeiro passo, na análise de séries temporais de contagens, é testar para correlação serial (autocorrelação). Uma forma é estimar um modelo de regressão aos dados, como por exemplo, a Regressão de Poisson, obter os resíduos e testar para correlação nula entre os *lags* correntes e dos resíduos (CAMERON e TRIVEDI, 2003).

Os estimadores usuais da média e das funções de autocovariância e autocorrelação amostrais (Equações 2.16, 2.17 e 2.18) são considerados fortemente consistentes para séries de contagens (DU e LI, 1991; GAUTHIER e LATOUR, 1994) sendo aplicados por Latour (1998); Silva e Silva (2002); Silva (2005); Cui e Lund (2009).

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t; \quad (2.16)$$

$$\hat{R}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (2.17)$$

e

$$\hat{\rho}(k) = \frac{\hat{R}(k)}{\hat{R}(0)}, 0 \leq k \leq N - 1, \quad (2.18)$$

onde X_t corresponde a contagens de itens não conformes no instante t (neste trabalho), N ao número de observações (amostras), \hat{R} é função de autocovariância amostral da observação de ordem k e $\hat{\rho}$ é a função de autocorrelação amostral da observação de ordem k.

Testes para independência também são alternativas para verificar a presença de autocorrelação. Vários testes, entre eles, testes de sequências (*runs test*) e uma adaptação do teste de Portmanteou, são usados por Jung e Tremayne (2006) para verificar a presença de correlação serial em dados de contagens.

Diversos testes foram desenvolvidos para verificar a independência das observações, com características voltadas a distribuições contínuas. Para dados de contagens, em que a distribuição não é discreta, o teste de aleatoriedade de Wald-Wolfowitz (1943) pode ser uma opção para complementar a análise, desde que os momentos sejam finitos e a variância maior que zero. É um teste baseado na correlação serial em que a hipótese nula é a de série ser aleatória.

Para finalizar outra opção é transformar os dados e proceder a análise como para variáveis contínuas com as funções de autocorrelação e autocorrelação parcial amostrais. Todavia, nenhum princípio matemático garante que uma transformação feita com determinado objetivo traga uma melhoria ao conjunto de dados sob outros aspectos; podendo até piorá-los. Quando a variável aleatória é expressa como uma contagem existe a possibilidade de observar um grande número de zeros, além de não assumir valores negativos, o que pode restringir a utilização de modelos gaussianos (ROSA, 1997).

Se os dados apresentarem autocorrelação, duas abordagens, de modo geral, têm sido aplicadas em diversos estudos. Uma corresponde a aplicar gráficos de controle com limites alargados. A outra envolve ajustar os dados a um modelo de séries temporais e monitorar seus resíduos (REYNOLDS e STOUMBOS, 2000).

Para dados discretos alguns modelos para captar a correlação serial são brevemente apresentados em seqüência. Broadbent (1958) propôs o uso de um modelo de dependência serial markoviano para capturar a correlação entre itens num ambiente de manufatura. Este modelo assume que o estado do item corrente depende apenas do estado do item anterior. Assim para uma cadeia de Markov com dois estados, duas probabilidades apenas são consideradas: a probabilidade de obter um item não conforme se o precedente for conforme e a probabilidade de obter um item conforme se o precedente for não conforme. Trabalhos correlatos podem ser encontrados em Bhat e Lal (1988); Bhat *et al.* (1990) e Lai, Govindaraju e Xie (2000).

Outra proposta é aquela sugerida por Lai, Govindaraju e Xie (1998). Neste caso, os autores sugerem o uso do modelo Binomial Generalizado, ou equivalentemente Binomial Correlacionado de Madsen (1993), para monitorar o processo. Sob este modelo a estatística "número de itens conformes testados até a aparição do primeiro ítem não conforme" teria uma distribuição geométrica correlacionada. Lai, Govindaraju e Xie (1998) mostraram os resultados de um estudo sobre o erro do tipo II (ou seja, diz que o processo "está sob controle" quando, na realidade, "não está") em situações nas quais o modelo probabilístico Binomial Generalizado é usado para o tratamento do número de itens não conformes do processo. A correlação entre as unidades amostrais supostamente uma constante igual para todas as unidades e o estimador proposto para o coeficiente de correlação é o de Madsen (1993).

A diferença entre estas duas metodologias é apontada por Mingoti e Carvalho (2003). No caso de Cadeias de Markov presume-se a injeção serial das unidades do processo, portanto, é preciso manter a

informação sobre o estado de cada item (isto é, se é "não conforme" ou "conforme"), na sequência exata em que foi inspecionado. Já no modelo Binomial Generalizado a informação necessária é aquela relacionada às amostras de itens do processo inspecionadas, sendo que em cada amostra apenas o número de "não conformes" precisa ser guardado.

Russo (2002) aplicou para variáveis não conformes (distribuição de Poisson) correlacionadas o modelo de regressão de Poisson. Para um mesmo processo estudou separadamente as variáveis contínuas (explicativas) e discretas (independentes), unindo-as depois por meio de uma função de transferência, aplicada aos resíduos.

Outras abordagens, envolvendo gráficos CUSUM para monitorar dados autocorrelacionados, são citadas no final do Capítulo 3.

2.9 SÍNTESE DO CAPÍTULO

Neste capítulo foi realizada uma revisão bibliográfica sobre os principais conceitos de Controle Estatístico de Qualidade e Gráficos Shewhart para atributos, especificamente o gráfico para não conformidades "np". Além dos limites de controle de Shewhart, especial atenção foi dada aos limites exatos (ou probabilísticos) e às suposições necessárias para a aplicação dos gráficos Shewhart.

3. GRÁFICOS DE CONTROLE DE SOMA ACUMULADA

Neste capítulo é feita a revisão bibliográfica sobre gráficos de controle de somas acumuladas para distribuição binomial. Inicialmente, na Seção 3.1, apresenta-se uma visão geral dos gráficos de controle de soma acumulada (CUSUM). Na Seção 3.2 são descritos os gráficos CUSUM para atributos com distribuição binomial. A Seção 3.3 traz as medidas de desempenho do CUSUM binomial, aprofundando-se na aproximação CD (“Corrected Diffusion”) de Reynolds e Stoumbos (1998, 1999, 2001).

3.1 GRÁFICOS DE SOMA CUMULATIVA (CUSUM)

Uma desvantagem de qualquer gráfico de controle do tipo Shewhart é que ele usa apenas a informação sobre o processo contida no último ponto plotado, ignorando qualquer informação dada pela seqüência inteira de pontos. Esta característica torna este gráfico relativamente insensível a pequenas mudanças no processo, da ordem de $1,5 \sigma$ (desvios padrão) ou menos (MONTGOMERY, 2004).

Além disso, o gráfico do tipo Shewhart é conhecido na literatura como lento na detecção de alterações para valores altos de ARL (SAMOHYL, 2009). Como visto no Capítulo 2, a incorporação de regras suplementares pode melhorar a sensibilidade do gráfico CUSUM. No entanto, o gráfico de controle mais apropriado para reconhecer o histórico dos dados é o de somas acumuladas CUSUM. É de uma riqueza e sofisticação conceitual e, sua aplicação não é difícil (SAMOHYL, 2009).

Os gráficos de controle de soma acumulada (CUSUM) podem ser considerados um aprimoramento dos gráficos do tipo Shewhart, e levam em conta a história dos dados, característica ausente nos gráficos mais simples. São capazes de reconhecer pequenas alterações nos processos muito antes dos alarmes dos gráficos Shewhart (SAMOHYL, 2009). São, portanto, uma alternativa eficaz ao gráfico de controle de Shewhart quando há interesse na detecção de pequenas a moderadas (e persistentes) mudanças (ALVES, 2003; MONTGOMERY, 2004). É um gráfico de controle avançado, desenvolvido para situações específicas, mas que ao mesmo tempo minimiza a ocorrência de alarmes falsos e alarmes não dados (SAMOHYL, 2009).

A idéia do gráfico de soma cumulativa é particularmente simples: considerando-se um gráfico de soma acumulada para a média de um processo, se o processo permanece sob controle no valor alvo, a soma acumulada é um passeio aleatório com média zero. Por exemplo, suponha o monitoramento da média de um processo, para amostras com $n \geq 1$, com μ_0 o valor alvo da média do processo e \bar{x}_j a média da j -ésima amostra. O gráfico de controle CUSUM é obtido plotando-se a quantidade (Equação 3.1) em relação à amostra i .

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0). \quad (3.1)$$

Se a média se desloca, então uma tendência, para cima, ou para baixo, de acordo com o deslocamento, se desenvolve na soma acumulada. Portanto, se os pontos plotados, evidenciam uma tendência, pode-se considerar que a média do processo mudou, sendo necessário pesquisar alguma causa atribuível (MONTGOMERY, 2004).

Em termos mais simples, para o cálculo a Equação 3.1 fica

$$C_i = C_{i-1} + (\bar{x}_i - \mu_0). \quad (3.2)$$

Uma série de somas acumuladas C_i negativas significa que as leituras observadas são inferiores aos valores nominais, mostrando um deslocamento da média do processo para baixo. Da mesma forma uma série de somas acumuladas C_i positivas significa que as leituras observadas são superiores aos valores nominais, mostrando um deslocamento da média do processo para cima. Ainda não é um gráfico de controle, pois faltam os limites de controle estatístico.

Um CUSUM trabalha acumulando desvios da média μ_0 que estão acima ou abaixo do alvo, com as estatísticas C_i^+ e C_i^- , respectivamente. Estas estatísticas são chamadas de CUSUM unilateral superior e inferior, respectivamente, e são calculadas de acordo com as Equações 3.3 e 3.4 (MONTGOMERY, 2004).

$$C_i^+ = \max[0, C_i^+ + x_i - (\mu_0 + k)] \quad (3.3)$$

e

$$C_i^- = \min[0, C_i^- + x_i - (\mu_0 - k)], \quad (3.4)$$

onde os valores iniciais são $C_0^+ = C_0^- = 0$.

Nas Equações 3.3 e 3.4, a constante k é usualmente chamada de valor de referência ou folga. A Equação 3.1 é sensível a qualquer desvio do valor alvo, mesmo de pequena magnitude. Todos os processos têm algum grau de variabilidade e eventualmente variações muito pequenas não são preocupantes. Logo, esta “tolerância” com pequenas modificações é formalizada explicitamente nas Equações do CUSUM com o valor de referência k . Este valor é diminuído dos desvios positivos e acrescentado aos desvios negativos. Em gráficos CUSUM para monitoramento da média, um valor comum, tanto na literatura como na prática, é 0,5 e significa que está se deixando de se preocupar com variações de meio desvio padrão no processo. O valor de referência k ajuda a melhorar o ARL_1 do processo, pois permite que o gráfico seja planejado para detectar alterações específicas, o que não ocorre com o procedimento de Shewhart (SAMOHYL, 2009).

Se tanto C_i^+ ou C_i^- excederem o valor de um intervalo de decisão h , o processo é considerado fora de controle estatístico. O intervalo de decisão é $[0, h]$ para C_i^+ , e $[-h, 0]$ para C_i^- .

Os valores iniciais são C_0^+ e C_0^- podem assumir valores diferentes de zero. Esse procedimento é chamado de RIR (Resposta Inicial Rápida) e foi elaborado para melhorar a sensibilidade do CUSUM no início do processo (LUCAS e CROISIER, 1982). Um valor comum é metade do intervalo de decisão h , ou seja, $C_0^+ = C_0^- = h/2$. Se o processo já inicia fora de controle estatístico, o sinal será dado mais rapidamente. Entretanto, caso o processo se inicie sob controle, este procedimento (RIR) terá pouco efeito (ROCHA, 2004).

Até aqui foi focalizado o gráfico CUSUM bilateral, que corresponde a dois procedimentos unilaterais, C_i^+ e C_i^- . Há situações que apenas um procedimento CUSUM unilateral é necessário, como por exemplo, monitorar apenas o aumento de não conformidades em uma linha de produção ou o aumento de casos de uma doença em uma região. A Figura 3.1 traz um exemplo de um gráfico CUSUM bilateral e a Figura 3.2 um gráfico CUSUM unilateral superior. Nestes gráficos é possível visualizar quando o processo se encontrou fora de controle estatístico (ultrapassou as linhas de controle) e se voltou ao estado de controle. Na Figura 3.2 pode-se ver que, mesmo antes de ultrapassar a linha de controle superior, o CUSUM indica que há uma tendência de aumento na média (ou outra estatística alvo) do processo.

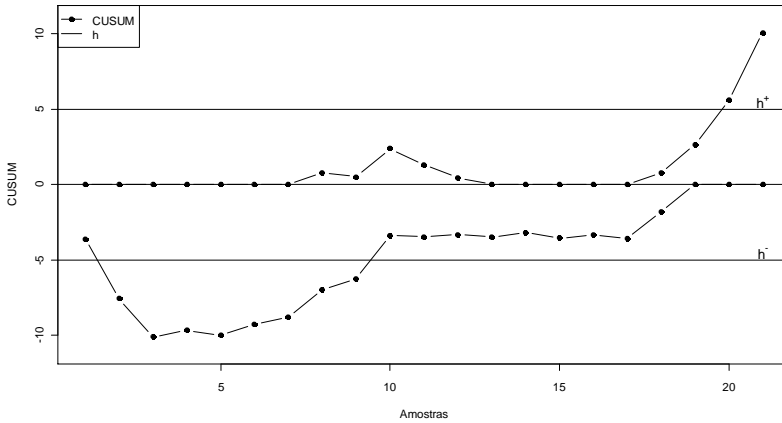


Figura 3. 1– Exemplo de um CUSUM bilateral.

Fonte: Adaptado de Scrucca (2004).

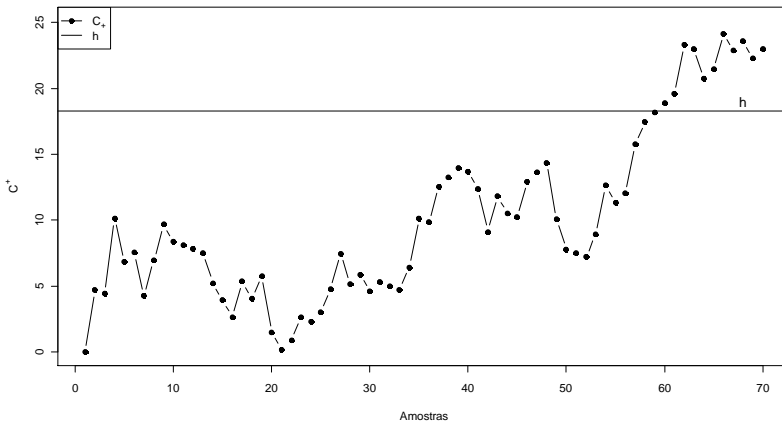


Figura 3.2 – Exemplo de um CUSUM unilateral superior.

Fonte: Adaptado de Gan (1993).

3.1.1 Avaliação de desempenho de gráficos CUSUM

No Capítulo 2 foi visto o RL, e seus momentos, entre eles o ARL, como medida de desempenho do gráfico de controle de Shewhart. Neste

capítulo, onde se aborda o gráfico CUSUM, é natural uma comparação dessas medidas. É através delas que é verificado que um procedimento é mais eficaz que outro, para uma situação específica.

Os gráficos de controle podem ser classificados pelas suas taxas de alarmes falsos, o tempo que se espera que ele sinalize um alarme falso, considerando que esteja sob controle estatístico. Entre todos os procedimentos com a mesma taxa de alarmes falsos, o procedimento considerado ótimo seria aquele que mais rapidamente detecta uma mudança. De um modo mais formal, entre todos os procedimentos com o mesmo ARL_0 , o procedimento ótimo será aquele com o menor valor esperado de tempo para sinalizar uma mudança, quando esta realmente ocorre (menor ARL_1) (HAWKINS e OLWELL, 1998).

Moustakides (1986) provou que o esquema CUSUM é ótimo nesse sentido, para uma particular mudança - pequena. Para todos os testes com o mesmo ARL_0 , o CUSUM obteve o menor RL esperado para o estado fora de controle, detectando bem mudanças de um desvio-padrão (1σ) na média. Não teve o mesmo desempenho para mudanças maiores.

Embora o CUSUM seja adequado para o diagnóstico de processos de uma particular alternativa o qual foi delineado, para Hawkins e Olwell (1998) alguns procedimentos CUSUM tem uma performance considerada robusta, não existindo um ponto, e sim um intervalo onde o CUSUM possa ser ótimo.

O gráfico CUSUM pode se tornar mais poderoso ao elevar a quantidade $\bar{x}_t - \mu$ por um expoente $w > 1$ na formulação do CUSUM. Valores maiores de w são indicados para detecção de mudanças maiores e vice-versa (JIAO e HELO, 2006; WU; JIAO; LIU, 2008).

A próxima seção aborda a fundamentação teórica do CUSUM. O enfoque estatístico dado é importante para a compreensão dos parâmetros do CUSUM.

3.2 CUSUM E SPRT

Historicamente existem duas correntes que convergem para a fundamentação teórica do CUSUM. A primeira, desenvolvida por Wald (1945), abrange uma seqüência de testes probabilísticos, denominados SPRT (*Sequential Probability Ratio Test*), de modo que estes são aplicados aos dados em seqüência, adicionando a informação obtida a cada nova observação disponível. Assumindo uma seqüência de

observações independentes X_i de tamanho n e deseja-se decidir entre as hipóteses nula (H_0) e alternativa (H_1). Associada a cada hipótese, está a densidade de probabilidade $f_0(x)$ e $f_1(x)$ respectivamente, para as medidas X_n . O conceito básico desta teoria é testar se os pontos têm distribuição f_0 ou f_1 . Alguns detalhes sobre os testes SPRT serão necessários e úteis para o procedimento CUSUM.

Outra linha é a desenvolvida por Page (1954), que propôs um CUSUM considerando desvios de um valor de referência, e sugeriu que o diagnóstico poderia ser feito observando-se o máximo afastamento de um ponto do CUSUM para outro (HAWKINS e OLWELL, 1998). O leitor verá que estas duas linhas, de Page (1954) e Wald (1945), estão de certo modo unidas.

Nos testes sequenciais (SPRT) é formada uma razão de verossimilhança Λ_n , sumarizada na Equação 3.5 (HAWKINS e OLWELL, 1998).

$$\Lambda_n = \prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)}. \quad (3.5)$$

O teste aceita H_0 se Λ_n for inferior ou igual a um determinado valor de corte A , e rejeita-se H_0 se Λ_n for superior a outra constante de corte B . Além disso, se $A < \Lambda_n < B$, então o teste remonta para outra observação X_{n+1} e atualiza a razão de máxima verossimilhança incorporando este resultado. Usualmente, é mais fácil trabalhar com o logaritmo da razão de máxima verossimilhança (Equação 3.6). Observa-se que $\ln \Lambda_n$ pode ser escrito como $\sum_{i=1}^n Z_i$ resultando na Equação 3.7 (HAWKINS e OLWELL, 1998).

$$\ln \Lambda_n = \sum_{i=1}^n \ln \frac{f_1(X_i)}{f_0(X_i)}. \quad (3.6)$$

e

$$Z_i = \ln \frac{f_1(X_i)}{f_0(X_i)}. \quad (3.7)$$

A variável escore Z_i é justamente a transformação da quantidade aleatória X_i . Como X_i é supostamente independente então Z_i também é. Portanto a regra de decisão do teste é sumarizada por:

i) aceita H_0 se $\ln \Lambda_n \leq \ln A$;

- ii) rejeita H_0 se $\ln \Lambda_n \geq \ln B$;
- iii) busca nova observação se $\ln A < \ln \Lambda_n \leq \ln B$.

Na prática, este teste funciona com a função log-verossimilhança, então o teste é baseado na soma acumulada de Z_i . A decisão de aceitar, rejeitar, ou continuar a amostragem é baseada no valor desta soma acumulada. Se a hipótese H_1 for verdadeira, então Z_i tem valores positivos. Na medida em que mais pontos são adicionados, a soma acumulada tende a subir. Se a hipótese H_0 for verdadeira, então Z_i tem valores negativos, e a soma acumulada tende a descer. As constantes A e B determinam as probabilidades de erro tipo I e erro tipo II. A decisão sobre os erros fixa os valores de A e B (HAWKINS e OLWELL, 1998).

As regras do CUSUM proposto por Page (1954) resultam em uma seqüência de testes de Wald (PAGE, 1954; HAWKINS e OLWELL, 1998). Um CUSUM unilateral pode ser pensado como uma repetição sucessiva deste teste. A diferença fundamental entre este CUSUM e o SPRT reside no fato de que o CUSUM não “encerra” quando a hipótese H_0 é aceita. Não se decide conceber que o processo está sob controle e interromper o processo de amostragem. Pelo contrário, o teste é reiniciado cada vez que há evidência de que se está sob controle. As evidências favorecem a hipótese nula sempre que a soma acumulada é negativa². Quando isto acontece, reinicia-se a soma a partir de zero.

Na abordagem CUSUM, é feita a amostragem até rejeitar-se a hipótese nula em favor da hipótese alternativa. A função escore Z_i pode ser considerada a mesma vista no SPRT, sendo uma função de X_i , dependendo dos parâmetros que as distribuições $f_0(X_i)$ e $f_1(X_i)$, sob e fora de controle estatístico, possam ter (HAWKINS e OLWELL, 1998).

Deste modo resume-se o problema que se busca resolver com os gráficos de controle: de determinar quando o processo muda de uma distribuição conhecida (“sob controle”) para uma segunda distribuição conhecida (“fora de controle”). Esta mudança ocorre em um instante desconhecido de tempo t.

3.3 GRÁFICOS CUSUM PARA ATRIBUTOS

Gráficos de soma acumulada para dados discretos, ou atributos, vão de encontro ao enfoque filosófico propagado pela teoria da Gestão

²Para o caso de um CUSUM unilateral superior.

da Qualidade de “fazer certo já na primeira vez”. O CUSUM também está de acordo com a filosofia de pontuar áreas que necessitam de particular atenção (“management by exception”). Por suas características próprias o gráfico CUSUM sinaliza quando há uma mudança no processo, além de permitir a indicação de quando um item não conforme (ou defeituoso) foi produzido (LUCAS, 1985).

Iniciando com Page (1954), existem diversos trabalhos abrangendo gráficos CUSUM para atributos: Bissel (1969); Ewan e Kemp (1960); Kemp (1962); Lucas (1985, 1989); Bourke (1991, 2001), Gan (1993); Hawkins e Olwell (1998), entre outros.

Um gráfico CUSUM para atributos (variáveis discretas) tem os mesmos conceitos que um CUSUM para variáveis mensuráveis. Algumas particularidades (valor nominal e parâmetros) são decorrentes das distribuições subjacentes à variável de interesse. Deste modo os gráficos são caracterizados de acordo com a distribuição da variável, sendo os mais comuns:

a) CUSUM de Poisson: para os casos de variáveis com distribuição de Poisson (não conformidades, contagens num intervalo de tempo);

b) CUSUM binomial: para os casos de variáveis com distribuição binomial (itens não conformes, proporções).

Existem procedimentos CUSUM para outras distribuições como Bernoulli (REYNOLDS e STOUMBOS, 1999), Geométrica (CHANG e GAN, 2001) e Binomial Negativa (RADAELLI, 1994).

Em se tratando de desempenho, os gráficos CUSUM para atributos são considerados mais sensíveis que os gráficos de Shewhart para pequenas e moderadas (e persistentes) alterações nos parâmetros de interesse (WOODALL, 1997). Mesmo assim o gráfico Shewhart “np” (ou seu equivalente “p”) ainda é considerado o mais utilizado do gráfico para atributos (WU e LUO, 2003). Algumas razões apontadas para não utilização dos gráficos CUSUM são:

a) Seu desenho e implementação não estarem tanto difundidos na comunidade científica (LUCAS, 1985);

b) Seu uso e interpretação são mais difíceis para os operadores;

c) Interesse no envolvimento dos operadores na melhoria da qualidade (BOX *et al.*, 1997).

Gráficos de controle de soma acumulada são mais comuns para controle de variáveis (contínuas) do que para atributos. E, infelizmente, mesmo para variáveis mensuráveis, o gráfico CUSUM, ainda, está ausente dos processos industriais (SAMOHYL, 2009).

3.4 CUSUM PARA VARIÁVEIS COM DISTRIBUIÇÃO BINOMIAL

Nesta seção serão descritos os gráficos de soma acumulada para atributos com distribuição binomial. Conceitualmente um gráfico CUSUM binomial examina o número de não conformidades acumuladamente numa seqüência de amostras. Pode-se dizer que é um do tipo “np”. Tem como objetivo a detecção quer de aumentos ou diminuições do número de não-conformidades “np”, como a maioria dos gráficos para atributos (WOODALL, 1997; HAWKINS e OLWELL, 1998; SANIGA; DAVIS; MCWILLIANS, 1995; BOURKE, 2001; WU; YEO; SPEDING, 2001; KHOO, 2004; WU; LUO; ZHANG, 2006; WU; JIAO; LIU, 2008).

É fácil de enxergar neste gráfico a base do conceito do CUSUM (SAMOHYL, 2009), resultando na soma acumulada dos desvios das observações ao redor da média (np) ou valor nominal do processo como ponto de referência.

Assim, um gráfico CUSUM binomial unilateral superior é aplicado para detectar um aumento no valor esperado de itens não conformes, do valor nominal np_0 para np_1 , e s corresponde ao acréscimo (%) em p_0 .

Seja uma série de amostras Y_i com $i \in \mathbb{Z}^+$, independentes, de tamanho n , com distribuição binomial. O gráfico de controle CUSUM unilateral superior é obtido plotando-se num gráfico a estatística C_i (Equação 3.8),

$$C_i = \max(0, C_{i-1} + X_i - k), \quad (3.8)$$

em relação a amostra Y_i , onde

- a) C_i é a estatística CUSUM com $C_0 = u, 0 \leq u \leq h$;
- b) X_i é o número de itens não conformes na amostra Y_i com $i = 1, 2, \dots$;
- c) k é uma constante denominada de valor de referência do CUSUM e depende da magnitude da mudança que se deseja detectar;

d) h é limite superior de controle do CUSUM.

Se o processo está sob controle, as somas acumuladas definidas em (3.8) descrevem um percurso aleatório com média zero. Porém, se a média muda para algum valor acima, então uma tendência ascendente se revelará na soma acumulada C_i . Se C_i ultrapassa o limite superior de controle h , diz-se que o processo está fora de controle estatístico. A Figura 3.2 ilustra um CUSUM binomial unilateral.

3.5 PARÂMETROS DE UM CUSUM BINOMIAL

Um CUSUM depende da escolha dos valores de k , h e consequentemente ARL_0 , considerando o estado inicial. Isto é possível utilizando tabelas, gráficos ou softwares. Um CUSUM completo envolve duas partes distintas. Primeiro a seleção de k e h para determinar um ARL_0 . A segunda parte envolve avaliar o quanto rapidamente o CUSUM (em controle) detecta uma mudança no parâmetro de interesse (HAWKINS e OLWELL, 1998). Em resumo, a construção de um CUSUM, leva as seguintes situações:

1. Dados h e k , encontrar o ARL ;
2. dados o ARL e k , encontrar h .

A situação definida em (2) será a abordada na maior parte deste trabalho. Significa que está se planejando um gráfico CUSUM para detectar uma mudança pré-especificada na proporção sob controle, com um valor tolerável mínimo para o ARL_0 (ou taxa máxima de alarmes falsos). O limite superior do CUSUM será determinado para estes valores de k e ARL_0 .

O ARL_0 sob controle do CUSUM depende então dos valores de h e k . Quanto maior a mudança desejada e o intervalo de decisão, maior o valor do ARL_0 . Alguns autores recomendam que seja preferível que o valor de h seja um número inteiro, para facilitar o planejamento e construção do gráfico. Esta restrição nos valores para o intervalo de decisão h implica num limitado campo de valores possíveis para o ARL_0 (HAWKINS e OLWELL, 1998; REYNOLDS e STOUMBOS, 1999, 2000). Como X é discreto pode ser encontrado um valor de h para um determinado ARL_0 alvo, todavia com algumas ressalvas. Supondo X um

valor inteiro, que segue uma distribuição discreta e k também é um inteiro, então apenas valores positivos para o CUSUM são possíveis. Isto significa que o ARL_0 será uma função do tipo escada de h , em que o acréscimo para h corresponde a valores inteiros análogos a partes fracionais das mudanças de h . Similarmente, se X for um inteiro e k um valor correspondendo a sua metade, os valores de C_n podem ser inteiros ou partes (metades) deste (HAWKINS e OLWELL, 1998).

É importante frisar que escolhendo k e um valor alvo de ARL_0 , o valor de h pode não atender exatamente ao alvo (ARL_0) desejado. É recomendável então selecionar um valor de h que corresponde ao valor mínimo para atender o ARL_0 alvo GAN (1993); (HAWKINS e OLWELL, 1998).

Nas próximas seções são abordados os parâmetros k , h e ARL de um CUSUM binomial.

3.6 VALOR DE REFERÊNCIA k DO CUSUM BINOMIAL

Nesta seção será focalizado o valor de referência k para um CUSUM Binomial. O objetivo é entender os procedimentos para escolha deste parâmetro.

A escolha do valor de referência k implica que o gráfico CUSUM será delineado para detectar uma particular magnitude de mudança desejada. Existem, para Hawkins e Olwell (1998), duas formas apropriadas para a escolha de k . A primeira refere-se circunstâncias específicas, onde existe um particular nível, se o processo está fora de controle, determinado previamente pelas próprias características do processo ou produto. A partir de cálculos teóricos ou experimentais é possível definir estes valores e o CUSUM pode ser delineado para detectar uma particular mudança nestes valores. A outra forma, e mais comum, é escolher o tamanho da mudança que se deseja detectar rapidamente. Isto implica que, existe uma mudança que é grande o suficiente para ter impacto sobre o processo, todavia pequena demais para ser detectada visualmente (HAWKINS e OLWELL, 1998).

Em se tratando de variáveis mensuráveis, k é sempre escolhido a meio caminho entre o valor alvo μ_0 e o valor da média (fora de controle) μ_1 , que se está interessado em detectar rapidamente. Sendo a mudança expressa em unidades de desvio padrão, então k corresponde à metade da magnitude da mudança (MONTGOMERY, 2004), sendo este k

considerado ótimo para a detecção de μ_1 (EWAN e KEMP, 1960; JOHNSON e LEONE, 1962).

O parâmetro k , de gráficos CUSUM para contagens, é determinado pela taxa de contagem aceitável e a taxa de contagem que se deseja detectar facilmente. Neste sentido, para um esquema CUSUM para variáveis com distribuição binomial seriam a média sob controle (np_0) e a média fora de controle estatístico (np_1), respectivamente.

Um primeiro procedimento para a escolha dos parâmetros de um CUSUM binomial foi baseado na aproximação à distribuição de Poisson. Este procedimento, de acordo com foi reproduzido em livros texto da época como em onde são encontradas Tabelas auxiliares para determinar os valores de k, h e n (KEMP, 1962; DUNCAN, 1986; BOURKE, 2001).

Posteriormente, alguns autores assumem que o valor de referência k para o CUSUM binomial pode ser escolhido usando a representação do gráfico CUSUM como uma seqüência de testes SPRT (HAWKINS, 1992, 1998; GAN, 1993; REYNOLDS e STOUMBOS, 1998). Uma primeira aproximação do valor de referência com o mesmo k do SPRT foi feita por Lucas (1985), para um CUSUM com distribuição de Poisson. Este valor de k é próximo ao nível da média que contém a menor variabilidade, deferindo da recomendação de Bissel (1969) de selecionar o valor de k quando a variância muda com uma alteração no nível da média. Esta recomendação (BISEL, 1969) serve para a distribuição normal, mas não para Poisson, em virtude de sua assimetria.

A escolha de parâmetros ótimos para um CUSUM binomial foi estudada por Gan (1993), que investigou a escolha do valor de referência, confirmando que o k do SPRT quantifica o desempenho do CUSUM, em termos de ARL. Até então, um desenho ótimo para procedimentos CUSUM para dados com distribuição binomial não existia na literatura da área.

Para facilitar a compreensão do trabalho de Gan, é necessário voltar às variáveis mensuráveis e um CUSUM para monitorar a média. Gan (1991) mediante extensivo trabalho numérico comprovou que, de fato, o k_{SPRT} é o mesmo k ótimo do CUSUM para detectar μ_1 . Neste sentido voltou sua atenção para o CUSUM binomial. Não era claro se as propriedades ótimas do SPRT poderiam servir para o CUSUM binomial. Então estudou o CUSUM binomial com o k_{SPRT} como valor para k , e analisou os resultados. Escolhido um valor adequado para k , o gráfico CUSUM é capaz de acumular a diferença $X_i - k$ com o valor

esperado de $E(X_i) - k$ ou $n(p_0 + \Delta p) - k$ e eventualmente conduzir a um sinal quando $C_i > h$. Para uma particular mudança Δp , tamanho de amostra n e um valor pré-fixado de ARL_0 , um gráfico CUSUM com o menor valor de ARL_1 entre outros gráficos de controle com o mesmo valor de ARL_0 é considerado um gráfico ótimo.

Gan (1993) propôs um valor ótimo para k correspondente a uma particular mudança Δp para um valor fixo de ARL_0 , partindo da relação existente entre o CUSUM e o SPRT.

Suponha que se tenha a mesma hipótese $H_0: \theta = \theta_0$ contra a hipótese $H_1: \theta = \theta_1$, onde $\theta_1 > \theta_0$. O SPRT emprega a razão de verossimilhança,

$$\Lambda_i = \frac{\prod_{j=1}^i P(X=x_j; p_0)}{\prod_{j=1}^i P(X=x_j; p_1)}, \quad n = 1, 2, \dots, \quad (3.9)$$

com $\theta_1 = p_1$ e $\theta_0 = p_0$. E, de forma equivalente,

$$\ln(\Lambda_i) = \sum_{j=1}^i (X_j - k_{SPRT}) = \ln(\Lambda_{i-1}) + X_i - k_{SPRT}, \quad (3.10)$$

onde o valor de referência (k_{SPRT}) é dado por

$$k_{SPRT} = \frac{n \ln((1-p_0)/(1-p_1))}{\ln((1-p_0)/(1-p_1)) - \ln(p_0/p_1)}. \quad (3.11)$$

Gan (1993) provou então que, também para um CUSUM binomial, o k_{SPRT} pode ser considerado o valor ótimo para o k do gráfico CUSUM, mediante resultados computacionais de um trabalho de comparações numéricas utilizando diversos valores de n , p_0 , p_1 e ARL_0 . Os resultados são consistentes com os resultados teóricos de Moustakides (1986) sobre o CUSUM ser ótimo para detecção de mudanças específicas. Assim este k é ótimo em termos de ARL como medida de desempenho (MORAIS, 2002). A dedução da Equação (3.11) (REYNOLDS e STOUMBOS, 1998) pode ser vista no AnexoA.

Hawkins (1992) e Hawkins e Olwell (1998) reforçam a ligação entre o k do CUSUM e do SPRT, desenvolvendo uma expressão aplicável a distribuições contínuas e discretas pertencentes à família exponencial. Se uma seqüência de variáveis aleatórias y seguindo uma distribuição membro da família exponencial na forma

$$f(y; \theta) = e^{a(y)b(\theta)+c(y)+d(\theta)}, \quad (3.12)$$

com função densidade de probabilidades f_{dp} para o caso contínuo e função de probabilidades f_p para o caso discreto, e fazendo Y_i igual a “quantidade suficiente” $Y_i = a(Y_i)$. Suponha que se tenha o mesmo teste de hipóteses anterior: $H_0: \theta = \theta_0$ contra a hipótese $H_1: \theta = \theta_1$, onde $\theta_1 > \theta_0$, onde θ é o parâmetro de interesse (média ou proporção, por exemplo). O SPRT é baseado na soma acumulada da quantidade $(Y_i - k)$ onde o valor de referência k é dado por

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)}, \quad (3.13)$$

e segue a forma de f_{dp} e f_p e a escolha de θ_0 e θ_1 . Deste modo é possível definir facilmente k , para variáveis com distribuições pertencentes a membros da família exponencial. Então, para a distribuição binomial:

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} = -n \frac{\ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}, \quad (3.14)$$

Onde,

$$f(y|m, \pi) = e^{y \ln\left(\frac{p}{1-p}\right) + n \ln(1-p) + \ln n! - \ln(y!(n-y)!)}, \quad (3.15)$$

com $d(\theta) = d(p) = n \ln(1-p)$ e $b(\theta) = b(p) = \ln\left(\frac{p}{1-p}\right)$.

A Fórmula (3.14) é a mesma da Equação (3.11). Assim, com as Equações 3.12 e 3.13 é possível deduzir o valor de k para distribuições pertencentes a família exponencial. Alguns exemplos estão no Anexo B.

Até então apenas foi abordado um CUSUM que inicia do zero (0). Outro valor inicial (RIR), não afeta a escolha de k . O valor de referência k do SPRT, utilizado neste trabalho como valor de referência k do CUSUM binomial, é determinado pela escolha dos dois valores de parâmetros p_0 e p_1 associados às hipóteses nula e alternativa.

3.6.1 Outras considerações sobre k

O valor de k é escolhido para uma resposta ótima a uma mudança de uma magnitude específica. Isto implica que não será ótimo para

mudanças cujo tamanho se afaste substancialmente da mudança desejada. Um gráfico planejado para detectar mudanças num patamar de três desvios padrão, pode não ser adequado para a detecção de alterações a um desvio padrão. Esta escolha vai estar diretamente relacionada com a natureza do processo que se deseja controlar. É possível elaborar procedimentos CUSUM que permitem detectar mudanças bem pequenas, como um quarto de um desvio padrão. E também mudanças maiores (HAWKINS e OLWELL, 1998).

A Expressão 3.11 não é a única para determinar o valor de referência k . Outras formas de determinar k aparecem na literatura. Uma forma análoga para o valor de referência $k = (\mu_0 + \mu_1)/2$ com $\mu_0 = np_0$, $\mu_1 = np_1$, com aproximação normal à binomial, resultando em $k = n/2$ foi aplicada por Schneider e O'Conneide (1987). Wu *et al.* (2008) apresentam um modelo que identifica valores ótimos de k e h , para um valor especificado de ARL_0 , minimizando o ARL_1 para $p = p_1$. O valor de referência k é tratado como uma variável independente. Este modelo é para um CUSUM binomial para detecção de grandes alterações ($p_1 \geq 2p_0$).

O arredondamento do valor de k é recomendado por alguns autores. São duas as razões e ambas levam em conta a conveniência (HAWKINS e OLWELL, 1998; REYNOLDS e STOUMBOS 1999, 2000). Como os valores do intervalo de decisão são zero (0) ou outro valor (h), em termos de k , com o arredondamento isto pode ser simplificado. A outra razão é em termos do cálculo do ARL considerando as cadeias de Markov, que será abordada na Seção 3.6.1.

3.7 O ARL E LIMITE SUPERIOR DE CONTROLE

Como exposto anteriormente, para determinar o limite superior de controle (h) do CUSUM, necessita que se especifique ARL_0 , ou vice-versa, para um dado valor de h , calcula-se o ARL_0 resultante. Assim, como é preciso escolher um deles para calcular o outro, optou-se por abranger toda esta teoria em conjunto.

A determinação das medidas de desempenho de um gráfico de controle, entre elas o ARL_0 , é fundamental para a avaliação de sua eficiência. Assim um gráfico é mais eficiente que outro se para um mesmo valor de ARL_0 ele tiver um valor menor de ARL_1 , ou seja, detecta mais rapidamente uma alteração quando esta realmente existe.

Uma das razões para os procedimentos CUSUM não serem de fácil implementação reside na caracterização do respectivo desempenho não ser necessariamente trivial (HAWKINS e OLWELL, 1998).

As abordagens para o cálculo das medidas de desempenho de um CUSUM binomial, que serão descritas a seguir. Nenhum destes métodos é simples, se comparado o aplicado nos gráficos de Shewhart, mesmo com limites probabilísticos. Os conceitos matemáticos são sofisticados e recursos computacionais são necessários. Inicialmente, são apresentados os métodos de Cadeias de Markov e simulação. Uma terceira linha para avaliar algumas medidas de desempenho está baseada nas aproximações desenvolvidas por Reynolds e Stoumbos (1999, 2000). Esta última, por ser a utilizada neste trabalho para o desenvolvimento de uma aproximação do limite superior h , será descrita com mais profundidade.

3.7.1 Uso de Cadeias de Markov para calcular as medidas de desempenho

Uma abordagem para determinar o ARL de um gráfico CUSUM para variáveis discretas é mediante cadeias de Markov. Esta abordagem originalmente foi proposta por Brook e Evans (1972). Permite determinar além do ARL_0 , os demais momentos e percentis da distribuição. Esse método pode ser utilizado para distribuições discretas, como uma aproximação acurada (BROOK E EVANS, 1972) (MORAIS, 2002) ou (considerada) exata (REYNOLDS e STOUMBOS, 2000).

Alguns autores aplicam a abordagem markoviana para procedimentos CUSUM binomiais, como Hawkins (1992); Gan (1993); Reynolds e Stoumbos (2000); Morais (2002); Lucas *et al.* (2006); Wu; Jiao; Liu (2008).

O esquema CUSUM possui características sumárias dependentes e em virtude do caráter recursivo das mesmas, estas podem ser vistas como constituindo uma cadeia de Markov em tempo discreto com espaço de estados discreto, uma vez que se está lidando com dados discretos MORAIS (2004).

O método é baseado na matriz de transição de probabilidades P , que representa a probabilidade de passar de um estado i para um outro estado j onde os diferentes estados representam diferentes acúmulos CUSUM (BROOK e EVANS, 1972; LUCAS, 1982; LUCAS e CROISIER, 1982).

A matriz transição de probabilidades P assume a seguinte forma,

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} & \mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix} \quad (3.16)$$

onde

i. \mathbf{R} representa vários estados sob controle com diferentes valores do CUSUM;

ii. \mathbf{p} é um vetor que representa a probabilidade de alcançar de um sinal fora de controle estatístico.

A matriz \mathbf{R} é denotada por (3.17) onde $F_\theta(i) = F_{bin(n,p_0+\theta)}(i)$ e $P_\theta(i) = P_{bin(n,p_0+\theta)}(i)$ representam a distribuição acumulada e de probabilidades de $Y_n(\theta) \forall i \in \mathbb{N}$.

O número de colunas da matriz \mathbf{R} está associada ao limite superior de controle h . Assim para um CUSUM com $k = 3$ e $h = 2$, a matriz \mathbf{R} terá 3 linhas e 3 colunas. O vetor \mathbf{p} , que representa a probabilidade de alcançar um sinal fora de controle estatístico é obtido calculando-se $1 - F_\theta(i)$. A matriz \mathbf{P} , portanto fica na forma exposta em (3.18).

O ARL é determinado resolvendo $\mathbf{u} = [\mathbf{I} - \mathbf{R}]^{-1} \mathbf{1}$ para qualquer tamanho da matriz \mathbf{R} . O primeiro elemento do vetor corresponde ao ARL_0 . A abordagem markoviana permite que o cálculo exato quando r_1/r_2 resultarem num número racional (REYNOLDS e STOUMBOS, 2000).

Todavia, o sucesso e acurácia dependem em muito do número de estados que tem a cadeia de Markov. Para um CUSUM binomial, esse número é o valor atingível no intervalo $[0, h]$. Limitando o número de valores atingíveis acelera-se o processo do cálculo do ARL. Por exemplo, se $h = 6$ e k for um inteiro, então a estatística CUSUM tem apenas 5 valores possíveis: 0, 1, 2, 3, 4 e 5. Se for usado $h = 5,771$ com quatro dígitos significativos tem-se cerca de 6000 possibilidades de valores para o cálculo do ARL, o que se torna computacionalmente complexo. Pelas duas razões, o valor teórico ótimo de referência $h = 5,771$, na prática é arredondado para 5,5; 5,75 ou 5. Qualquer uma dessas escolhas pode resultar em modificações no ARL_0 do que o inicialmente pretendido (HAWKINS e OLWELL, 1998).

$$\mathbf{R} = \begin{bmatrix} F_\theta(k) & P_\theta(k+1) & P_\theta(k+2) & \cdots & P_\theta(k+(h-1)) \\ F_\theta(k-1) & P_\theta(k) & P_\theta(k+1) & \cdots & P_\theta(k+(h-2)) \\ F_\theta(k-2) & P_\theta(k-1) & P_\theta(k) & \cdots & P_\theta(k+(h-3)), \\ F_\theta(k-3) & P_\theta(k-2) & P_\theta(k-1) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \\ F_\theta(k+1-h) & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad (3.17)$$

e

$$\mathbf{P} = \begin{bmatrix} F_\theta(k) & P_\theta(k+1) & P_\theta(k+2) & \cdots & P_\theta(k+(h-1)) & 1 - F_\theta(k+(h-1)) \\ F_\theta(k-1) & P_\theta(k) & P_\theta(k+1) & \cdots & P_\theta(k+(h-2)) & 1 - F_\theta(k+(h-2)) \\ F_\theta(k-2) & P_\theta(k-1) & P_\theta(k) & \cdots & P_\theta(k+(h-3)), & 1 - F_\theta(k+(h-3)) \\ F_\theta(k-3) & P_\theta(k-2) & P_\theta(k-1) & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ F_\theta(k+1-h) & \vdots & \vdots & \vdots & \ddots & 1 - F_\theta(k) \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.18)$$

Por isso, Hawkins e Olwell (1998) argumentam que seja preferível que o valor de h seja um número inteiro, para facilitar o planejamento e construção do gráfico. Esta restrição nos valores para o intervalo de decisão h implica num limitado campo de valores possíveis para o ARL (HAWKINS e OLWELL, 1998; REYNOLDS e STOUMBOS, 1999, 2000). Além disso, quando p_0 é muito pequeno, a matriz de transição de probabilidades da cadeia de Markov pode ser muito grande, resultando em “grande esforço computacional” (REYNOLDS e STOUMBOS, 2000).

Hawkins (1992) propôs um algoritmo para determinar o ARL de esquemas CUSUM para distribuições contínuas e discretas pertencentes à família exponencial, refinando o modelo proposto por Brook e Evans (1972), mas obtendo os mesmos resultados. O caminho inverso, determinar h , para um pré-determinado valor de ARL_0 , é um pouco mais complexo. Höhle (2007), baseado no algoritmo de Hawkins (1992), complementa-o com um processo iterativo que gera valores para h , calcula o ARL_0 e aplica interpolação logarítmica até atingir uma precisão pré-especificada.

3.8 TÉCNICAS DE SIMULAÇÃO

Outra abordagem para determinar o ARL é simular processos produzindo o CUSUM, guardar os valores de RL obtidos e então calcular a média destes. A coleta de valores de RL também pode ser obtida para derivar uma distribuição empírica destes.

Quando se utiliza este método, gera-se uma sequência de observações de uma distribuição pré-determinada e calcula-se o valor da estatística de interesse até que esta fique fora dos limites de controle estatístico. O número médio de observações necessárias até a emissão de um sinal, isto é, o valor do RL, varia de uma sequência para outra. A estimativa para o ARL obtém-se repetindo o procedimento um elevado número de vezes (10000 ou mais). Para facilitar, determina-se o ARL assumindo que as mudanças no valor médio em relação ao valor pretendido ocorrem no instante em que inicia o procedimento (ALVES, 2009).

Esta técnica pode ser computacionalmente trabalhosa e não produzir os resultados almejados no caso de processos com valores maiores de ARL (HAWKINS e OLWELL, 1998). Técnicas de redução

da variância são sugeridas para melhorar a estimação de valores de ARL por simulação (JUN e CHOI, 1993; ROSS, 1990).

Para gráficos CUSUM binomiais, não há trabalhos em que a simulação é usada para determinar especificamente parâmetros de desempenho.

A simulação pode ser útil para checar a implementação das cadeias de Markov ou outros métodos de cálculo de ARL (HAWKINS e OLWELL, 1998). Técnicas de simulação são usadas neste trabalho para avaliar parâmetros e medidas de desempenho.

3.8 APROXIMAÇÃO CD (“CORRECTED DIFFUSION”)

Uma terceira linha para avaliar algumas medidas de desempenho está baseada nas aproximações desenvolvidas inicialmente por Wald (1947) e na extensão deste trabalho proposta pela teoria de correção de difusão desenvolvida por Siegmund (1979, 1985). Reynolds e Stoumbos (1999, 2001) desenvolveram aproximações para o cálculo de algumas medidas de desempenho, embasadas nestes conceitos. Essa abordagem foi denominada de “Corrected Diffusion Approximation” e, neste trabalho, chamada de aproximação CD. Essas aproximações compõem uma forma relativamente simples de determinação dos parâmetros de um CUSUM de Bernoulli e Binomial para aplicações práticas.

3.9.1 A aproximação de Siegmund

Wald (1947) desenvolveu originalmente o SPRT (Seção 3.2) e derivou algumas aproximações simples para as curvas características de operação e o tempo médio até a sinalização de uma mudança. Essas aproximações ignoram o “excesso” da estatística SPRT sobre os limites de aceitação ou rejeição ao fim do teste e não são consideradas muito acuradas. Quase quatro décadas depois, para algumas distribuições pertencentes à família exponencial Siegmund (1979, 1985) desenvolveu uma teoria baseada na “correção difusa” ou “corrected diffusion” (Correção CD) que aproxima o limite do valor esperado do excesso da estatística teste do SPRT sobre as linhas de aceitação e rejeição corrigindo este excesso na aproximação de Wald. O princípio desta aproximação foi substituir a condição de passeio aleatório da estatística de controle por um processo de movimento browniano contínuo.

Então para o caso de variáveis aleatórias membros da família da distribuição exponencial de um parâmetro, foi possível obter aproximações para o valor de ARL de um CUSUM. As observações devem ser independentes e identicamente distribuídas, de acordo com uma pré-fixada distribuição pertencente à família exponencial. Para a distribuição normal (escrita na forma da família exponencial) com média θ e variância igual a um (1), Siegmund (1985) aproxima o ARL por,

$$ARL \cong \frac{|e^{(-2\theta b)} + 2\theta b - 1|}{2\theta^2}, \quad (3.19)$$

onde θ é a média do processo e b corresponde ao limite superior do CUSUM.

Todavia os resultados da Equação 3.19 não foram precisos quando comparados com os obtidos numericamente por Van Dobben de Bruyn (1968) *apud* Siegmund (1985). Siegmund (1985) aprimora esta Equação, corrigindo o “excesso” da estatística SPRT sobre os limites de aceitação ou rejeição. Ele aproxima o limite do valor esperado deste excesso. Assim, inclui-se a aproximação anteriormente citada que faz com que $e(b) = e(b + 2 * 0,583) = e(b + 1,166)$. Para facilitar a compreensão será feita uma modificação na nomenclatura utilizada daqui para diante. O limite de decisão será chamado de h , como é comum na literatura, e o limite corrigido de b . Portanto a expressão é sumarizada em 3.20, na forma conhecida na literatura como “Aproximação de Siegmund”.

$$ARL \cong \frac{e^{-2\theta(h+1,166)} + 2\theta(h+1,166) - 1}{2\theta^2}. \quad (3.20)$$

Na Expressão 3.20 o limite de decisão h foi corrigido por $b = h + \rho_+ - \rho_-$, onde ρ corresponde à aproximação do excesso sobre o intervalo de decisão h do CUSUM. Siegmund (1985) e Wu (2005) colocam que $\rho_{\pm} = E_0 R_{\pm\infty}$ e

$$\rho^{(a)}(\theta) = E_{\theta} \left(R_{\infty}^{(a)} \right), \quad \text{para } a > 0, \quad (3.21)$$

onde R_{∞} é uma variável aleatória da distribuição para o resíduo entre a estatística calculada e a regra de decisão h . E, $\rho^{(a)}$ corresponde aos (a) momentos dessa distribuição.

Siegmund (1985) apresenta uma aproximação para ρ_+ , que pode ser facilmente resolvida por métodos numéricos computacionais. A distribuição deve estar na forma da família exponencial com um parâmetro, com $P_\theta(z_n \in dx) = dF_\theta(x) = e^{\theta z(x) - \psi(\theta)} dF(x)$. Assumindo que a distribuição é contínua e fortemente não aritmética então ρ_+ pode ser aproximado por,

$$\rho_+ = \frac{\kappa}{6} - \int_0^\infty \frac{1}{\pi \lambda^2} \operatorname{Re} \left(\log \left(2 \left(\frac{1-g(\lambda)}{\lambda^2} \right) \right) \right), \quad (3.22)$$

com $g(\lambda) = e^{\psi(i\lambda)}$, e $\kappa = \psi^{(3)}(0)$ corresponde ao terceiro cumulante de $F(x)$. Na Equação (3.22) λ é uma variável aleatória que está associada ao intervalo de confiança do valor estimado onde há mudança da média do processo de um estado sob controle estatístico para um estado considerado fora de controle estatístico. No Anexo C estão mais informações sobre a aproximação de Siegmund. No entanto, ao leitor interessado em aprofundar-se neste assunto, recomenda-se a leitura de Siegmund (1985).

3.9.2 Aproximação CD para distribuições de Bernoulli e Binomial

A teoria exposta na seção anterior só é válida para distribuições fortemente não aritméticas (SIEGMUND, 1985; WU, 2005). Para Johnson; Kemp e Kotz (2005, p. 74) a maioria das distribuições discretas utilizadas pertence à outra classe menos abrangente denominada de distribuições aritméticas. De acordo com Siegmund (1985) uma variável aleatória X é denominada aritmética se para uma constante $d > 0$, $P\{X \in \{\dots, -2d, -d, 0, d, 2d, \dots\}\} = 1$.

Reynolds e Stoumbos (1998, 1999) colocam que a distribuição de Bernoulli também não atende condições apontadas por Siegmund (1985). Portanto, inicialmente os autores concluíram que esta abordagem não poderia ser aplicada ao SPRT para o monitoramento da proporção p em dados com distribuição de Bernoulli. Nesse sentido, a distribuição binomial também não atende aos pressupostos básicos que devem ser considerados para a aproximação de Siegmund.

No entanto, Reynolds e Stoumbos (1998) conseguiram estender a abordagem de Siegmund (1979, 1985) no ARL de um SPRT para monitorar a proporção p . A variável foi padronizada e aproximou-se esta variável transformada a um processo de movimento browniano.

Então um argumento de ajuste de curvas foi usado para derivar esse excesso da estatística teste do SPRT além de h . O desenvolvimento da aproximação CD considerando uma variável de Bernoulli para p é delineado a seguir, de acordo com o exposto em Reynolds e Stoumbos (1998, 1999).

Reynolds e Stoumbos (1999, 2001) desenvolveram uma aproximação para o ARL de um gráfico CUSUM, denominada daqui para diante de aproximação CD, para variáveis de Bernoulli empregando um valor ajustado do limite de decisão h , denotado por b , numa fórmula relativamente simples. O valor ajustado de h é dado por,

$$b = h + \varepsilon(p_0)\sqrt{p_0(1-p_0)}, \quad (3.23)$$

onde $\varepsilon(p_0) \approx \rho_+$, que corresponde a aproximação do excesso sobre o intervalo de decisão do CUSUM.

Para a dedução, considerou-se X uma variável aleatória de Bernoulli com $P(X = 1) = p_0$ onde $0 < p_0 < 1$, e efetuou-se a transformação padronizada da variável X de modo a obter,

$$Z = \frac{X - p_0}{\sqrt{p_0(1-p_0)}}, \quad (3.24)$$

O terceiro cumulante κ é dado por

$$\kappa = E_{p_0}(Z^3) = \left(\frac{1-p_0}{p_0}\right)^{1/2} - \left(\frac{1-p_0}{p_0}\right)^{-1/2} \quad (3.25)$$

e

$$g(\lambda) = E_{p_0}(e^{i\lambda Z}), \quad (3.26)$$

Então

$$E_{p_0}(e^{i\lambda Z}) = e^{\left(i\lambda\left(\frac{1-p_0}{p_0}\right)^{1/2}\right)}p_0 + e^{\left(-i\lambda\left(\frac{1-p_0}{p_0}\right)^{-1/2}\right)}(1-p_0), \quad (3.27)$$

onde $i = \sqrt{-1}$. Chamando $\rho_+ = \varepsilon(p_0)$, substituindo-se em 3.22, fica-se com

$$\rho_+ = \frac{1}{6}E_{p_0}(Z^3) - \frac{1}{\pi} \lim_{s \rightarrow 0} \int_0^\infty \frac{1}{\lambda^2} \operatorname{Re} \log \left\{ \frac{2}{\lambda^2} \left[1 - E_{p_0}(e^{i\lambda Z}) e^{-\frac{(\lambda s)}{2}} \right] \right\} d\lambda. \quad (3.28)$$

A Equação 3.28 foi desenvolvida por Siegmund (1985), sendo apropriada para obter numericamente ρ_+ (SIEGMUND, 1985; WU, 2005). Na Equação 3.28, λ é uma variável aleatória, e está associada ao intervalo de confiança da estimativa da média do processo onde há mudança de um

Reinolds e Stoumbos (1998, 1999) definem que quando $p = p_0$ o limite do excesso esperado da estatística além de h é igual a $\varepsilon(p_0)\sqrt{p_0(1-p_0)}$, por meio de um argumento de ajuste de curvas. Além disso, $\varepsilon(p_0)$ satisfaz a seguinte relação,

$$\varepsilon(p_0) = -\frac{1}{3}E_{1-p_0}(Z^3) + \varepsilon(1-p_0) = \frac{1}{3}E_{p_0}(Z^3) + \varepsilon(1-p_0). \quad (3.29)$$

Para $0 < p_0 \leq 0,5$, uma aproximação simples para $\varepsilon(p_0)$ que produz bons resultados para p é

$$\varepsilon(p_0) \approx \frac{E_{p_0}(Z^3)}{3}. \quad (3.30)$$

Do mesmo modo uma aproximação para o excesso sobre h é:

$$E_{p_0}(Z^3) \frac{\sqrt{p_0(1-p_0)}}{3} = \frac{1-2p_0}{3}. \quad (3.31)$$

No intervalo $0,5 < p_0 \leq 1$, uma boa aproximação é obtida avaliando-se $\varepsilon(1-p_0)$ nas equações acima (REYNOLDS e STOUMBOS, 1998, 1999).

As aproximações CD para as medidas de desempenho do SPRT de Reynolds e Stoumbos (1998) foram obtidas utilizando-se as aproximações originais de Wald (1947), substituindo h por $b = h + \frac{1}{3}(1-2p_0)$ com $p_0 < p_1$. Reinolds e Stoumbos (1999) também estenderam essas aproximações para calcular o ARL de um gráfico CUSUM com distribuições de Bernoulli e binomial. Estas últimas aproximações, para variáveis binomiais, serão detalhadas na próxima seção.

3.9.3 Fórmulas para calcular o ARL com a correção de h

Reynolds e Stoumbos (2001) propuseram fórmulas para o cálculo de ARL_0 e ARL_1 para um gráfico CUSUM com variáveis binomiais. Estas fórmulas, conforme visto na Seção 3.9, tem embasamento nos trabalhos de Wald (1947) e Siegmund (1985). O cálculo do ARL_0 para um CUSUM binomial pode ser sumarizado por

$$ARL_0 \approx \left(\frac{r_2}{n(r_2 p_0 - r_1)} \right) \left(b - \frac{n\gamma(e^{br_2} - 1)}{1 - e^{-n\gamma r_2}} \right), \quad (3.32)$$

onde

$$r_1 = -\ln \frac{1-p_1}{1-p_0}; \quad r_2 = \ln \left(\frac{p_1(1-p_0)}{p_0(1-p_1)} \right) \text{ e}$$

$$\gamma = r_1/r_2. \quad (3.33)$$

Segundo Reynolds e Stoumbos (2001) a Equação (3.32) permite que se calcule diretamente o ARL_0 , uma vez conhecidos a proporção sob controle (p_0), a proporção considerada fora de controle, que se deseja detectar rapidamente (p_1), o tamanho da amostra (n) e o limite de decisão (h). Esta fórmula para o ARL_0 usa o valor corrigido do limite de decisão h , denotado por b , obtido com a aproximação 3.23. O valor ajustado de h , para $0 < p_0 < 0,5$ é dado por,

$$b = h + \frac{1}{3}(1 - 2p_0). \quad (3.34)$$

Pode se notar também que o valor de referência k corresponde a γ (3.33) multiplicado pelo tamanho da amostra n .

$$k = n\gamma. \quad (3.35)$$

Para o ARL_1 Reynolds e Stoumbos (2001) também apresentam uma fórmula que é

$$ARL_1 \approx \left(\frac{r_2}{n(r_2 p_1 - r_1)} \right) \left(b - \frac{n\gamma e^{-n\gamma r_2}(e^{br_2} - 1)}{e^{br_2}(1 - e^{-n\gamma r_2})} \right). \quad (3.36)$$

Reynolds e Stoumbos (2001) ainda recomendam que, por conveniência, γ seja um número racional. Isto implica que $\gamma = 1/l$ e

$r_2/r_1 = l$ onde l é um número inteiro positivo. Isso também estende-se a k . Assim tanto k quanto h serão múltiplos de l tornando, na opinião dos Reynolds e Stoumbos (2001), mais simples a construção do gráfico.

Para a construção de um CUSUM binomial os valores de p_0 e p_1 podem ser pré-especificados, e com esses valores determina-se o valor de referência k ($n\gamma$). Para as aproximações Reynolds e Stoumbos (2001) colocam ser necessário algumas vezes ajustar p_1 , de modo que $l = 1/\lambda$ seja um inteiro positivo. Resumindo, as fórmulas propostas por Reynolds e Stoumbos (2001) para o cálculo do ARL consideram o caso especial em que $p_0 < 0,5$ e $l = 1/\lambda$ seja um inteiro positivo. Extensões do método para casos mais gerais estavam, na época, em desenvolvimento.

O procedimento CUSUM binomial de Reynolds e Stoumbos (2000, 2001) é ligeiramente diferente do CUSUM tradicional (Equação 3.4). A estatística CUSUM permite valores negativos, pois $C_i = \max(0, C_{i-1}) + (X_i - n\gamma)$. A razão foi de que era necessário saber o valor real de C_i , quando este é considerado igual a zero, para determinar algumas propriedades estatísticas do CUSUM. Do ponto de vista prático, os dois procedimentos são equivalentes.

A seção seguinte trata da acurácia das fórmulas para ARL_0 e ARL_1 com a aproximação CD.

3.9.4 Acurácia

Como as fórmulas da aproximação CD têm origem nos trabalhos de Siegmund (1979, 1985) optou-se por relatar inicialmente a acurácia da aproximação de Siegmund. Para Hawkins e Olwell (1998) a aproximação de Siegmund para o ARL produz bons resultados para pequenos a moderados valores do parâmetro θ , todavia podendo ser pouco fiel para grandes valores. De acordo com Rogerson (2006) a aproximação de Siegmund é menos acurada para valores maiores de k , ou seja, magnitudes maiores, resultando que o tempo até a ocorrência de um alarme falso seja menor do que o esperado.

Reynolds e Stoumbos (1999) consideram que para um CUSUM com variáveis de Bernoulli, a aproximação CD foi considerada acurada para $0 < p \leq 0,5$. Para valores $0,5 < p < 1$ os Reynolds e Stoumbos (1999) avaliaram $\varepsilon(1 - p_0)$ obtendo também bons resultados.

Chang e Gan (2001) colocam que a aproximação CD para o mesmo CUSUM não é adequada para valores de $p < 0,01$. O erro

relativo pode ser muito grande pelo fato de que para valores muito pequenos de p a distribuição é fortemente assimétrica. Assim, a variável padronizada de Bernoulli (e binomial) não é bem aproximada pelo processo de movimento browniano.

3.10 CONSIDERAÇÕES FINAIS

As suposições do Capítulo 2 também se aplicam ao gráfico CUSUM binomial. Os dados devem ser verificados quanto à independência, presença de superdispersão e aderência à distribuição binomial. As recomendações também são às mesmas.

Particularmente relativo à aderência, o desempenho de um esquema CUSUM é muito sensível a afastamentos da distribuição, podendo alterar o cálculo do ARL. Também a presença de superdispersão (variância maior que $np(1-p)$) resulta também no aumento do número de alarmes falsos. Utilizar um CUSUM nestes dados iria subestimar o valor atual de alarmes falsos que este processo iria produzir (HAWKINS e OLWELL, 1998). Se os dados forem autocorrelacionados o gráfico CUSUM também irá sinalizar mais alarmes falsos, do que o esperado (MONTGOMERY, 2004).

As recomendações são as mesmas do Capítulo 2. Devem ser aplicadas técnicas que capturem a autocorrelação. Algumas propostas (de CUSUM) para dados com distribuição de Poisson autocorrelacionados estão nos trabalhos de Weiß (2007, 2009); Weiß e Testik (2009), que trazem alternativas interessantes, com aplicações dos gráficos EWMA, CUSUM e modelo INAR (First-Order Integer-Valued Autoregressive Process). Um gráfico de controle CUSUM “livre de distribuição” para dados autocorrelacionados foi proposto por Kim *et al.* (2007). Foi aplicado para uma série de contagens em que aderência à distribuição de Poisson, era afetada (negativamente) pelo alto valor do desvio padrão do processo (superdispersão).

Para finalizar o Capítulo, na seqüência está um quadro comparativo (Tabela 3.1) do gráfico Shewhart (Capítulo 2) e do procedimento CUSUM, adaptado de Rocha (2004). Enquanto que o gráfico Shewhart é efetivo na detecção de mudanças de grande magnitude e temporárias, o CUSUM é comumente considerado eficaz na sinalização de pequenas a moderadas (e persistentes) mudanças.

Tabela 3. 1 - Comparação entre gráficos de controle do tipo Shewhart e CUSUM.

	Shewhart	CUSUM
Causas	Esporádicas e grandes	Crônicas e pequenas a moderadas
Informação sobre a mudança	Só sinaliza	Estima o momento e tamanho da mudança
Teste de hipótese	Contínuo individual Rejeita ou não H_0	Sequencial Pede outra amostra ou rejeita H_0
Memória	Sem	Com
Implementação	Fácil	Complexa e trabalhosa

FONTE: Adaptado de Rocha (2004).

É conhecido na literatura que o gráfico do tipo Shewhart é lento com altos valores de ARL_0 para detectar alterações no processo (SAMOHYL, 2009). Em termos de processos industriais um valor comum é $ARL_0 = 370$. Todavia a escolha do valor de ARL_0 depende do que se deseja monitorar. Um exemplo, na área de epidemiologia, de Sanches (2000), para a construção do CUSUM para fixou o $ARL_0 = 1.000$, para dados semanais. Isso significa que o valor esperado de uma situação fora de controle estatístico, quando o sistema realmente está sob controle, deverá ocorrer uma vez a cada mil semanas, ou seja, aproximadamente, a cada 19,2 anos, uma exigência considerada bastante forte.

3.11 SÍNTESE DO CAPÍTULO

Neste capítulo descreveu-se o procedimento de soma acumulada CUSUM. Iniciou-se com conceitos gerais e, então a maior parte do Capítulo foi dedicada a fundamentação teórica dos gráficos CUSUM para variáveis com distribuição binomial. Abrangeu-se a ligação do CUSUM com o SPRT, o cálculo do valor de referência k , do limite superior h e do ARL, com destaque para a aproximação CD.

4. GRÁFICOS COMBINADOS SHEWHART-CUSUM

Este capítulo traz uma revisão bibliográfica e os principais conceitos associados aos gráficos combinados Shewhart-CUSUM. Inicialmente, abrange-se a fundamentação teórica e o histórico de aplicações para variáveis mensuráveis. Em seguida focaliza-se o gráfico combinado para atributos. Na seqüência, abrange-se uma abordagem para o cálculo das medidas de desempenho, incluindo alguns exemplos.

4.1 GRÁFICOS COMBINADOS SHEWHART-CUSUM

Nos Capítulos anteriores ficou claro que um gráfico CUSUM é indicado para sinalizar pequenas e moderadas alterações, enquanto que um gráfico do tipo Shewhart é mais efetivo para as alterações maiores. Todavia, se a magnitude da mudança é desconhecida ou se modifica ao longo do tempo, nenhum dos gráficos mencionados, planejados para pequenas ou grandes alterações, irá ter um bom desempenho para todas as situações. Um tratamento possível é combinar múltiplos gráficos para abranger mudanças de diversas magnitudes. Como exemplo, adicionar limites de Shewhart aos gráficos CUSUM ou EWMA para detectar tanto pequenas como grandes alterações (TSUNG e WANG, 2010).

Assim, uma abordagem para aumentar a sensibilidade do gráfico CUSUM, para grandes mudanças, é o procedimento combinado Shewhart-CUSUM (MONTGOMERY, 2004). Um gráfico combinado Shewhart-CUSUM incorpora num mesmo eixo os valores observados, os limites de controle Shewhart, a estatística CUSUM e o intervalo de decisão do CUSUM. É uma modificação simples do procedimento CUSUM, com a adição do gráfico de controle Shewhart.

Este gráfico pressupõe, de um modo geral, quatro linhas de controle: os limites superiores e inferiores de Shewhart (LCS e LCI) e o intervalo de decisão h (h^+ e h^-). Deve apresentar também os valores das estatísticas correspondentes às amostras (observadas ou padronizadas) e os valores das somas cumulativas (positiva e negativa). Em virtude disto a análise pode se tornar confusa (ROCHA, 2004). Um exemplo de gráfico combinado bilateral está na Figura 4.1.

Algumas reduções podem ser feitas de acordo com o caso a ser tratado. Por exemplo, se apenas é desejada a investigação do problema,

um gráfico unilateral superior pode ser suficiente (ROCHA, 2004). A Figura 4.2 ilustra um gráfico combinado Shewhart-CUSUM unilateral superior.

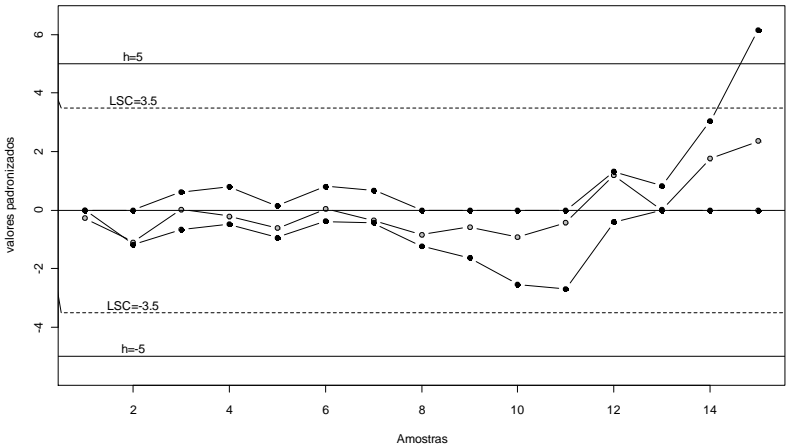


Figura 4. 1 – Exemplo de gráfico combinado bilateral.

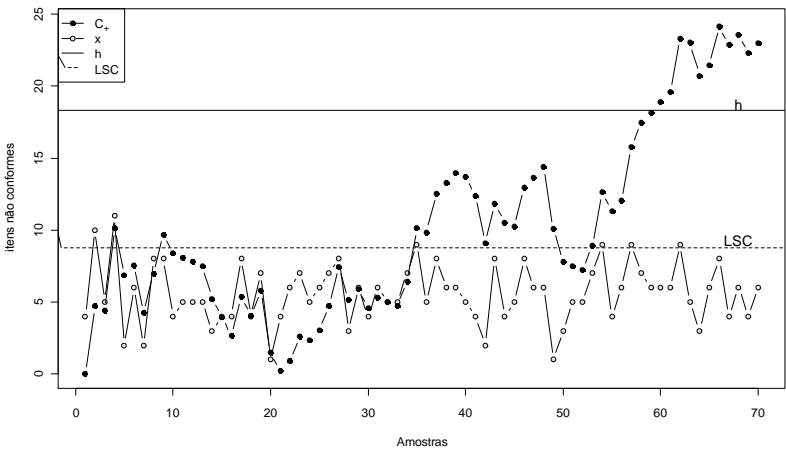


Figura 4. 2 – Exemplo de gráfico combinado unilateral superior.

Fonte: Adaptado de Moraes (2002)

A primeira citação de um gráfico de controle combinado foi feita no trabalho de Westgard *et al.* (1977), que aplicou o gráfico combinado como uma alternativa ao CUSUM, para tentar melhorar o controle em laboratórios de análises clínicas. A partir de simulação fez variar o tamanho dos subgrupos, a média e a variância do processo. Chegou à conclusão de que seria possível melhorar aquele sistema de controle de qualidade utilizando o esquema combinado CUSUM-Shewhart, pela maior probabilidade de detecção de desvios simulados. Sua proposta, no entanto não foi prontamente aceita neste meio (MORAIS, 2002).

O trabalho de Lucas (1982) trouxe algumas descobertas importantes do uso de um gráfico combinado bilateral. Lucas (1982) efetuou uma análise comparativa do comportamento desses três esquemas de gráficos de controle para média de processo com distribuição normal, Shewhart, CUSUM e Shewhart-CUSUM. Concluiu que o gráfico combinado Shewhart-CUSUM tem um valor de ARL_0 mais baixo que o CUSUM padrão para o mesmo valor de h , o que significa mais alarmes falsos. Como sugestão, para a prática, na aplicação de gráfico combinado, pode ser vantajoso um ligeiro aumento no valor de h , de modo que se obtenha um ARL_0 um pouco maior. Deste modo, com a adição das linhas de Shewhart para um CUSUM com $k = 0,5$ e $h = 4$ ou 5 , propôs limites LSC com $q = 3,5\sigma$ ou 4σ . Esses limites causam apenas um pequeno decréscimo no ARL_0 e ainda fornecem valores menores de ARL_1 , do que o CUSUM, para mudanças maiores que 2σ aproximadamente. Limites LSC superiores a 4σ não são indicados e, um LSC igual a 3σ implica numa grande redução do valor do ARL_0 . Assim, Lucas (1982) estudou um CUSUM com $k = 0,5; h = 5,2 (5,1)$ e $LSC = 3,5 (4,0) \sigma$, e concluiu que este apresenta desempenho similar ao CUSUM individual para pequenas alterações e sinaliza antes do CUSUM individual para mudanças maiores (2σ e $2,5\sigma$ respectivamente). Na Figura 4.3, visualiza-se o desempenho em termos de $\log(ARL)^3$ de um gráfico combinado (LUCAS, 1982) para monitoramento da média quando comparado com um CUSUM individual com os valores da Tabela 4.1.

O gráfico da Figura 4.3 traz comparações para os valores de ARL_0 do CUSUM e dos gráficos combinados. São 465, 508 e 476, respectivamente.

³É utilizado o logaritmo do ARL para facilitar a visualização.

Tabela 4. 1 - Valores de ARL para os seguintes gráficos: CUSUM ($k = 0,5$ e $h = 5$); combinado ($k = 0,5$; $h = 5,1$ e $LSC = 4,0$) e combinado ($k = 0,5$; $h = 5,2$ e $LSC = 3,5$).

Desvio do valor alvo (múltiplo de σ)	ARL		
	CUSUM $h=5$	COMBINADO $h=5,1$ $LSC=4,0$	COMBINADO $h=5,2$ $LSC=3,5$
0	465	508	476
0,25	139	148	149
0,50	38	39,2	39,9
0,75	17	17,4	17,6
1,00	10,4	10,6	10,6
1,50	5,75	5,84	5,81
2,00	4,01	4,06	3,94
2,50	3,11	3,11	2,91
3,00	2,57	2,47	2,22
4,00	2,01	1,60	1,38
5,00	1,69	1,16	1,07

Fonte: Adaptado de Lucas (1982:57).

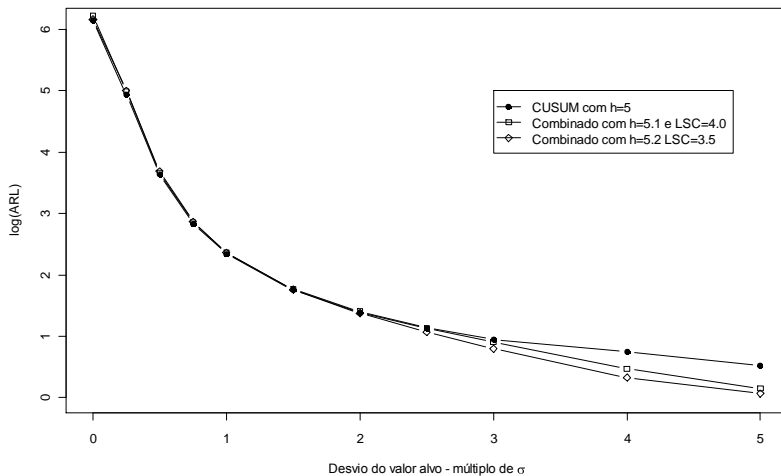


Figura 4. 3– Comparação de valores de $\log(ARL)$ da Tabela 4.1.

Fonte: adaptado de Lucas (1982).

Um gráfico similar ao proposto por Lucas (1982) adiciona os limites de Shewhart ao procedimento de escores acumulados (MUNFORD, 1980) no monitoramento da média de um processo. Este gráfico apresentou maior sensibilidade na detecção de pequenas e médias alterações (WOODALL e NCUBE, 1984).

Yashchin (1985) estudou gráfico combinados unilaterais e bilaterais, com variáveis contínuas normalmente distribuídas e discretas com distribuição de Poisson. Sua maior contribuição (MORAIS, 2002) é que o desenho do gráfico não é somente baseado no comportamento do ARL, e sim na função de “sobrevivência” do RL na situação sob controle.

No monitoramento ambiental está uma das mais consolidadas aplicações do gráfico combinado Shewhart-CUSUM. A Agência de Proteção Ambiental dos Estados Unidos utiliza o esquema combinado unilateral para monitorar a contaminação de águas subterrâneas (USEPA, 1989). E, ainda nesta área, alguns trabalhos analisam a aplicação e desempenho de gráficos combinados no monitoramento (de médias) de contaminação de aquíferos (GIBBONS, 1999; CHOU *et al.*, 2001; ZHOU *et al.*, 2008), com suas especificidades.

Outras aplicações do gráfico combinado para médias abrangem o monitoramento de processos laboratoriais (SOLBERG, 1984; BLACKSELL *et al.*, 1994, 1996; ROCHA, 2004; CONRATHS e SCHARES, 2006) e monitoramento dos erros de previsão (SOUZA e SAMOHYL, 2008).

Uma modificação no gráfico combinado para médias foi sugerida por Wu *et al.* (2008). Uma mudança no cálculo da quantidade CUSUM, conforme Equações 4.1, 4.2, 4.3 e 4.4, é feita com o objetivo de otimizar o seu desempenho. Esta mudança corresponde na aplicação de um expoente $w > 1$ na formulação do CUSUM (ver seção 3.1.1).

$$C_i^+ = \max(0, C_{i-1}^+ + q - k) \quad (4.1)$$

e

$$C_i^- = \min(0, C_{i-1}^- + q + k), \quad (4.2)$$

com

$$q = \begin{cases} z_t^w, & e_{z_t} \geq 0 \\ -(-z_t)^w, & e_{z_t} < 0, \end{cases} \quad (4.3)$$

onde

$$z = \frac{x - \mu_0}{\sigma}. \quad (4.4)$$

Este gráfico apresentou, de modo geral, desempenho superior ao CUSUM individual e o combinado Shewhart-CUSUM padrão numa escala de 0 a 6 desvios padrão.

4.2 GRÁFICOS COMBINADOS PARA ATRIBUTOS

A literatura apresenta poucos trabalhos sobre gráficos combinado para atributos. Resultados numéricos relativos ao gráfico combinado superior unilateral CUSUM-Shewhart, considerando dados com distribuição de Poisson podem ser encontrados em Yashchin (1985); Abel (1990) e Hawkins e Olwell (1998).

Um esquema combinado Shewhart-CUSUM para variáveis discretas com distribuição binomial, do tipo “np” foi abordado por Morais (2002) e Morais e Pacheco (2006), que estudaram a ordenação estocástica do RL. E, como estudo de caso, investigaram o comportamento do RL e seus momentos do gráfico combinado Shewhart-CUSUM binomial, com um planejado para detectar uma alteração de 114% (cerca de $1,6\sigma$) na proporção sob controle p_0 . Os parâmetros h e LSC foram fixados e os valores de ARL calculados por abordagem markoviana. Os parâmetros e valores de ARL_0 e ARL_1 estão na Tabela 4.2.

Tabela 4. 2–Valores de k , h , ARL_0 e ARL_1 para os gráficos CUSUM, Shewhart e combinados ($n=100$, $p_0=0,02$ e $p_1=0,0427685$).

Gráfico	ARL_0	ARL_1	k	h	LSC
CUSUM	1015,71	5,932	3	6	
SHEWHART	1073,03	7,815			7
COMBINADO	603,743	5,648	3	6	4

FONTE: Adaptado de Morais e Pacheco (2006).

Na Tabela 4.2 estão apenas os valores de ARL, mas Morais e Pacheco (2006) determinaram outros momentos e alguns percentis. O limite superior do CUSUM e do combinado são os mesmos e o limite de Shewhart se tornou mais estreito neste último. Efetuaram-se algumas variações nos parâmetros dos gráficos combinados: valor de h (para 7) e LSC (6 e 8), além de um CUSUM com RIR ($ARL_0 = 995,07$ e $ARL_1 = 3,99$). Morais (2002), ainda analisa um gráfico combinado Shewhart-CUSUM com RIR ($ARL_0 = 592,56$ e $ARL_1 = 3,92$). Morais e Pacheco (2006) concluem que o gráfico combinado sinaliza antes do CUSUM individual a mudança (p_1), mas também apresenta mais alarmes falsos, na parte Shewhart, aconselhando aplicar um CUSUM individual com RIR, ao invés de se acrescentar as linhas de Shewhart.

Ressaltam-se algumas diferenças desses trabalhos e do proposto neste trabalho. Enquanto que naqueles os autores fixam os limites e verificaram o que ocorre com o ARL, pretende-se fixar o ARL_0 combinado e efetuar o balanceamento dos limites de controle dos gráficos individuais. Morais e Pacheco (2006) analisam o desempenho até $p = p_1$ enquanto que neste trabalho será verificado o que ocorre quando $p < p_1$, $p = p_1$ e $p > p_1$.

Quanto a aplicações a dados reais, até a redação deste trabalho não se tem registros do uso de um gráfico combinado Shewhart-CUSUM binomial.

4.3 CÁLCULO DOS PARÂMETROS DO GRÁFICO COMBINADO

Esta seção descreve o cálculo dos parâmetros, mais especificamente o ARL e formas distintas de aproximação. A abordagem adotada neste trabalho é vista em seguida, com exemplos abrangendo variáveis com distribuição normal e binomial.

4.3.1 ARL de um gráfico combinado Shewhart CUSUM

Para Coelho (2008) as diversas aplicações de gráficos combinados Shewhart-CUSUM possuem suas próprias especificações, adequadas a cada caso (WESTGARD *et al.*, 1977; GIBBONS, 1999; ZHOU *et al.*, 2007; SOUZA e SAMOHYL, 2008). A abordagem markoviana foi utilizada por alguns autores para o cálculo das medidas

de desempenho do esquema combinado tanto de variáveis contínuas (LUCAS, 1982) como discretas (YASHCHIN, 1985; MORAIS, 2002; MORAIS e PACHECO, 2006). Um estudo com simulação foi realizado por Rocha (2004) que construiu tabelas de ARL em função dos parâmetros dos gráficos (limites de controle de Shewhart, k e h do CUSUM) para variáveis mensuráveis. O RL de esquemas combinados Shewhart-CUSUM foi investigado sob a ótica da ordenação estocástica por Morais (2004).

A combinação de gráficos de controle, considerando a função de probabilidade de ocorrência de alarmes falsos, foi abordada por Montgomery (2004); Souza e Samohyl (2008); Coelho (2008) e Wu *et al.* (2008).

De modo geral, se há r variáveis estatisticamente independentes para um produto em particular, e se o gráfico com probabilidade de erro tipo I é mantido para cada uma, então a verdadeira probabilidade (erro tipo I) para cada procedimento de controle conjunto, é $\alpha' = 1 - (1 - \alpha)^r$ onde $(1 - \alpha)^r$ é a probabilidade de que todas as variáveis sejam representadas no gráfico simultaneamente dentro de seus limites de controle (MONTGOMERY, 2004; ALVES, 2009). Seguindo este raciocínio, quando dois gráficos diferentes monitoram a mesma variável, a taxa de alarmes falsos do gráfico combinado α_{cs} será uma combinação de taxas individuais de cada um, α_s e α_c para Shewhart e CUSUM respectivamente. Assim, uma expressão que sumariza essa combinação é dada por

$$\alpha_{cs} = \alpha_s + \alpha_c - \alpha_s \alpha_c. \quad 4.5$$

No Exemplo 4.1 Coelho (2008) ilustra a aplicação da Equação 4.5.

Exemplo 4.1: Dois gráficos de controle com taxas de alarme falso de 1% individualmente geram um gráfico combinado com taxa de alarme falso de cerca de 2%.

$$\alpha_{cs} = \alpha_s + \alpha_c - \alpha_s \alpha_c = 0,01 + 0,01 - 0,01 * 0,01 = 0,0199 \cong 2\%$$

Portanto, desejando-se manter a taxa no valor original (1%), menor que 2%, os limites de controle dos dois gráficos devem ser recalculados, resultando em valores mais distantes da linha central (SOUZA e SAMOHYL, 2008).

Assim, no planejamento de gráfico é necessário ajustar o LSC (limite superior de Shewhart) e h (CUSUM) de modo a alocar o erro tipo I (taxa de alarme falso - α) do procedimento combinado. Se o LSC é relaxado, h necessita ser mais estreito, e o gráfico combinado terá um poder maior para detectar pequenas mudanças. Similarmente se o LSC for mais estreito, h será mais largo, tendo o gráfico um poder maior para detectar grandes mudanças (WU *et al.*, 2008).

Uma vez determinados os valores de ARL_0 do combinado e dos esquemas individuais, podem-se encontrar os parâmetros do CUSUM e limites de controle de Shewhart, pelo método que for mais conveniente ao pesquisador. Os exemplos 4.2 e 4.3 mostram como ficam os parâmetros do CUSUM e do gráfico do tipo Shewhart para dois gráficos combinados.

Exemplo 4.2⁴: adaptado de Coelho (2008). Deseja-se um gráfico combinado para monitoramento da média com $ARL_0 = 370$ (comum em processos industriais), equivalente ao do gráfico Shewhart isolado, com parâmetros do CUSUM iguais a $k = 0,5$ e $h = 5$, os limites de Shewhart devem ser ajustados a uma distância de 3,5 desvios padrão da média ($q_w = 3,5$). O valor do ARL_0 do gráfico combinado, obtido por simulação, conforme Rocha (2004), é igual a 397. Aplicando a Expressão 4.5, partindo-se do ARL_0 combinado desejado de 370 e CUSUM com parâmetros acima, então os valores aproximados de ARL_0 e taxa de alarme falso para os gráficos combinado, Shewhart e CUSUM são: ARL_0 do gráfico combinado = 370 ($\alpha = 0,27\%$); ARL_0 do gráfico CUSUM = 465,44 ($\alpha = 0,21\%$) e ARL_0 do gráfico tipo Shewhart = 1800,48 ($\alpha = 0,06\%$).

Exemplo 4.3: adaptado de Wu *et al.*(2008). Deseja-se um gráfico combinado para monitoramento da média com $ARL_0 = 300$, limites de Shewhart com $q_w = 3,75$ ($\alpha = 0,018\%$, $ARL_0 = 5655$), e CUSUM com valor de referência $k = 0,5$. Ao se aplicar a Expressão 4.5, partindo-se do ARL_0 combinado desejado de 300 com os dados anteriores, então os valores aproximados de ARL_0 e taxa de alarmes falsos da parte CUSUM são aproximadamente: ARL_0 CUSUM = 318 e $\alpha = 0,32\%$. O valor do limite superior do CUSUM é $h = 3,9^4$.

⁴Resolvido com auxílio da função `xcusum.crit` do pacote `spc` (KNOTH, 2009) que estima o valor de h (por cadeias de Markov) sendo dados k e ARL_0 de um CUSUM para variáveis com distribuição normal.

Neste trabalho, a Expressão 4.5 será aplicada para o caso dos gráficos combinados para atributos com distribuição binomial. A taxa de alarmes falsos de Shewhart será determinada conforme exposto no Capítulo 2. Os exemplos 4.4 e 4.5 ilustram o cálculo dos parâmetros dos gráficos para um esquema combinado binomial. O valor do ARL_0 do CUSUM e o parâmetro h são calculados usando-se o algoritmo de Hawkins (1992)⁵. O valor de k é calculado aplicando-se a Equação 3.11.

Exemplo 4.4: adaptado de Moraes (2002) e Gan (1993). Para amostras de tamanho $n = 100$, distribuição binomial, considerando que se deseja gráfico combinado que detecte rapidamente uma mudança de 0,05 a 0,056 (12% em p_0) e o ARL_0 do esquema combinado deve ser igual a 240. Um gráfico pode ser construído com os seguintes parâmetros:

- a) $k = 5,29$ e $h = 18,3$ para o CUSUM, resultando num ARL_0 de aproximadamente 250;
- b) $LSC = 13,19$ do gráfico Shewhart com $q_w = 3,76$ e ARL_0 aproximado de 6000.

Os valores das taxas de alarme falso dos esquemas isolados CUSUM e SHEWHART são 0,4% e 0,017%. Então, o ARL_0 do esquema combinado será de 240 com taxa de alarme falso de 0,42%.

Exemplo 4.5: adaptado de Moraes e Pacheco (2006). Neste exemplo considera-se $n = 100$, distribuição binomial, e deseja-se detectar uma mudança de 0,02 a 0,0427685 (114% em p_0) com:

- a) O ARL_0 do CUSUM é 1015,71;
- b) O ARL_0 de SHEWHART é 1073,03.

Um gráfico pode ser construído com os seguintes parâmetros:

- c) $k = 3$ e $h = 7$ para o CUSUM;
- d) $LSC = 6,62$ do gráfico Shewhart com $q = 3,3$.

O ARL_0 do esquema combinado será de aproximadamente 522 com taxa de alarme falso de 0,19%.

⁵Pacote surveillance (Höhle, 2007) GNU R

Para que a adição dos limites de Shewhart tenha ação efetiva na performance do gráfico CUSUM, a seguinte condição deve ser satisfeita: $LSC \leq h + k$ (LUCAS, 1982; YASHCHIN⁶, 1985; ROCHA, 2004).

Quando $LSC \leq k$, um sinal fora de controle estatístico só pode ser emitido apenas na parte Shewhart. Por este motivo os gráficos de controle de Shewhart clássicos podem ser considerados um caso especial do gráfico combinado CUSUM-Shewhart. O mesmo ocorre se k e h são iguais a zero. (YASHCHIN, 1985). Se $LSC > h + k$, não é um gráfico combinado (ROCHA, 2004).

Isto fica bem claro para gráfico de monitoramento da média. Para gráficos Shewhart-CUSUM voltados a dados discretos não há ainda um detalhamento no que se refere a k . O valor de referência k , especificamente para o gráfico CUSUM binomial, pode assumir valores grandes, inclusive para mudanças de pequena magnitude, pois depende do tamanho da amostra n . Para Morais (2002), $LSC < h$, caso contrário o desempenho do gráfico combinado é igual a do CUSUM unilateral.

4.4 SÍNTESE DO CAPÍTULO

Este Capítulo trouxe uma revisão bibliográfica e os principais conceitos associados aos gráficos combinado Shewhart-CUSUM. Abrangeu tópicos relativos ao embasamento teórico e aplicações, para variáveis mensuráveis e atributos. O cálculo das medidas de desempenho, com alguns exemplos, complementou este capítulo.

⁶Para Yashchin (1985): $LSC < h + k$.

5. APROXIMAÇÕES PARA O LIMITE SUPERIOR DE UM CUSUM BINOMIAL

O objetivo deste Capítulo é desenvolver um modelo matemático para obtenção do limite superior do CUSUM binomial. As fórmulas são derivadas das expressões para o cálculo de ARL_0 com a aproximação CD (“Corrected Diffusion”). O objetivo principal foi criar uma fórmula que permitisse o cálculo direto do limite de decisão “h”. Para atingir esse objetivo empregou-se a função LambertW e seu desenvolvimento em série. Assim, não há a necessidade do uso de métodos iterativos para encontrar raízes. É necessário ter definido o valor de proporção p_0 sob controle e fora de controle p_1 que se deseja rapidamente detectar, o tamanho da amostra n , além do ARL_0 desejado. Os valores obtidos foram comparados com resultados encontrados na literatura, com a obtenção de raízes por meio de um método numérico, pela abordagem markoviana e algumas simulações. Para finalizar, a fim de contornar algumas limitações da fórmula, desenvolveu-se um algoritmo em que a aproximação é integrada a um processo de interpolação logarítmica e à abordagem markoviana.

5.1 APROXIMAÇÃO DO LIMITE SUPERIOR DO CUSUM

Conforme visto no Capítulo 3, subseção 3.8.3, Reynolds e Stoumbos (2001) mostram que expressão para o cálculo do ARL_0 de CUSUM binomial é dada por

$$ARL_0 \approx \left(\frac{r_2}{n(r_2 p_0 - r_1)} \right) \left(b - \frac{n\gamma(e^{br_2} - 1)}{1 - e^{-nr_2}} \right), \quad (5.1)$$

onde

$$r_1 = -\ln \left(\frac{1-p_1}{1-p_0} \right); \quad (5.2)$$

$$r_2 = \ln \left(\frac{p_1(1-p_0)}{p_0(1-p_1)} \right); \quad (5.3)$$

$$\gamma = r_1/r_2 \quad (5.4)$$

e

$$b = h + (1 - 2p_0)/3. \quad (5.5)$$

O valor de referência k corresponde a (5.6), facilmente obtido por meio de operações algébricas.

$$k = n\gamma. \quad (5.6)$$

A Equação (5.1) permite que se obtenha diretamente o ARL_0 , uma vez conhecidos a proporção sob controle (p_0), a proporção considerada fora de controle, que se deseja detectar rapidamente (p_0), o tamanho da amostra (n) e o limite de decisão (h). A correção em h , definida por (5.5) depende do valor de p_0 .

É possível encontrar h utilizando a Expressão (5.1) para um pré-determinado valor de ARL_0 , mediante um método numérico para solução de Equações. Métodos iterativos, envolvendo ou não o uso derivadas, podem ser utilizados. Embora os recursos computacionais atuais tornem o uso de métodos numéricos simples, a escolha de valores iniciais adequados é necessária para a convergência destes métodos. Assim, procurou-se uma forma de aproximação que permitisse o cálculo direto do parâmetro h , sem a necessidade de empregar métodos iterativos.

Uma abordagem interessante é a de Rogerson (2006) que obteve uma fórmula direta para o cálculo de h de um CUSUM para monitoramento da média de um processo. Rogerson (2006) partiu da fórmula de aproximação de Siegmund (1985),

$$ARL \approx \frac{e^{-2\theta b} + 2\theta b - 1}{2\theta^2}, \quad (5.7)$$

onde $\theta = \delta - k$, com δ definido como a magnitude da mudança que se deseja detectar, k o valor de referência e $b = h + 1,166$ com h correspondendo ao valor crítico do limite de decisão do CUSUM. O parâmetro h é corrigido somando-se o valor de 1,166 de acordo com Siegmund (1979,1985). Rogerson (2006) aplicou a função LambertW, descrita por Corless *et al.* (1996) e sua avaliação numérica, descrita por Chapeau-Blondeau e Monir (2002), na resolução de (5.7) para b .

Para um dado argumento x , a função LambertW retorna os valores $W(x)$ que satisfazem a Expressão (5.8). Chapeau-Blondeau e

Monir (2002) denotam que se o argumento x é negativo e próximo de zero (0), a função pode ser expandida de acordo com (5.9).

$$W(x)e^{W(x)} = x \quad (5.8)$$

e

$$\begin{aligned} & LambertW\{-1, x\} = \\ & L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(L_2-2)}{2L_1^2} + \frac{L_2(6-9L_2+2L_2^2)}{6L_1^3}, \end{aligned} \quad (5.9)$$

com $L_1 = \ln(-x)$ e $L_2 = \ln(-\ln(-x))$.

Mediante o desenvolvimento com os primeiros três termos em (5.9), Rogerson (2006) propõe a Equação (5.10) para encontrar b . Esta equação é considerada acurada para valores de $k < 1$. Rogerson (2006), ainda, complementa que a acurácia dos resultados depende da quantidade $k^2 ARL_0$. Então, para a obtenção de bons resultados k necessita ser maior que $1/\sqrt{ARL_0}$.

$$b \approx \left(\frac{2k^2 ARL_0 + 2}{2k^2 ARL_0 - 1} \right) \left(\frac{\ln(1 + 2k^2 ARL_0)}{2k} \right). \quad (5.10)$$

Neste sentido, procurou-se aplicar a metodologia utilizada por Rogerson (2006) para determinar o valor do parâmetro h de um CUSUM binomial⁷. Desenvolveu-se a Equação (5.10) para b , resultando na expressão (5.11).

$$\begin{aligned} b \approx & \frac{1}{r_2} \left\{ - \left(\frac{nr_2\gamma}{e^{-nr_2\gamma} - 1} \right) + nARL_0(p_0r_2 - r_1) - \right. \\ & LambertW \left[- \left(\frac{1}{e^{nr_2\gamma} - 1} \right) \right. \\ & \left. \left. \left(e^{nr_2\gamma + n(r_1 - p_0r_2)} \left(-ARL_0 \frac{r_2\gamma}{(1 - e^{-nr_2\gamma})(r_1 - p_0r_2)} \right) nr_2\gamma \right) \right] \right\}. \end{aligned} \quad (5.11)$$

⁷Rogerson (2006) usou o pacote computacional MAPLE (Versão 8). Neste trabalho para a parte algébrica, foi utilizado o pacote Wolfram Mathematica 7.0 for Students, licença n°. L3283-1202, função ProductLog[].

A quantidade entre colchetes na Equação (5.11) foi desenvolvida considerando-se dois, três, quatro e cinco termos, conforme a Equação (5.9). Com a utilização de três termos em (5.11) e com b corrigido por (5.6), a Equação (5.10) se torna igual a

$$h \approx \frac{1}{r_2} \left(- \left(\frac{nr_2\gamma}{e^{-nr_2\gamma}-1} \right) + nARL_0(p_0r_2 - r_1) - \left[L_1 - L_2 + \frac{L_2}{L_1} \right] \right) - \frac{(1-2p_0)}{3}, \quad (5.12)$$

onde

$$L_1 = -\ln \left(- \left(\frac{1}{e^{nr_2\gamma} - 1} \right) \left(e^{nr_2\gamma + n(r_1 - p_0r_2)} \left(-ARL_0 - \frac{r_2\gamma}{(1 - e^{-nr_2\gamma})(r_1 - p_0r_2)} \right) nr_2\gamma \right) \right) \quad (5.13)$$

e

$$L_2 = \ln(-L_1). \quad (5.14)$$

A inclusão de mais termos resultaria teoricamente em uma aproximação mais acurada. Assim, as Expressões (5.15) e (5.16) contêm 4 e 5 termos em (5.11), com L_1 e L_2 definidos por (5.13) e (5.14)

$$h \approx \frac{1}{r_2} \left(- \left(\frac{nr_2\gamma}{e^{-nr_2\gamma}-1} \right) + nARL_0(p_0r_2 - r_1) - \left[L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(L_2-2)}{2L_1^2} \right] \right) - \frac{(1-2p_0)}{3}, \quad (5.15)$$

e

$$h \approx \frac{1}{r_2} \left(- \left(\frac{nr_2\gamma}{e^{-nr_2\gamma}-1} \right) + nARL_0(p_0r_2 - r_1) - \left[L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(L_2-2)}{2L_1^2} + \frac{L_2(6-9L_2+2L_2^2)}{6L_1^3} \right] \right) - \frac{(1-2p_0)}{3}. \quad (5.16)$$

Na seção seguinte é feita análise da acurácia (viés) das Equações (5.12), (5.15) e (5.16) quando comparadas com os valores obtidos por um método numérico iterativo.

5.2 ACURÁCIA DAS APROXIMAÇÕES

A acurácia (ou exatidão) das aproximações corresponde à proximidade dos resultados aos valores reais. Inicialmente, verificam-se quão próximos estão os valores se comparados com a solução da Equação (5.1) para b . Os valores obtidos para h com as Equações (5.12), (5.15) e (5.16), são comparados aos resultados da aplicação de um método numérico iterativo para a solução da Equação (5.1) para b , efetuando a correção dada por (5.6). Assim, o valor de h é aproximado resolvendo a Equação (5.17). Esta avaliação é necessária, pois as aproximações são alternativas à solução da Equação.

$$ARL_0 - \left(\frac{r_2}{n(r_2 p_0 - r_1)} \right) \left(\left(h + \frac{(1-2p_0)}{3} \right) - \frac{n\gamma \left(e^{\left(h + \frac{(1-2p_0)}{3} \right) r_2 - 1} \right)}{1 - e^{-n\gamma r_2}} \right) = 0. \quad (5.17)$$

Foram elaboradas duas rotinas no aplicativo GNU R para a análise proposta. A primeira (hrs.1) calcula h com as Equações (5.12), (5.15) e (5.16). A segunda (hrs) aproxima h com (5.17)⁸. As rotinas estão no Apêndice A e B respectivamente. Os resultados com apenas dois termos em (5.12) também foram verificados. Mas, como estes são muito pouco acurados, optou-se por focar apenas as Equações com três ou mais termos. As rotinas foram aplicadas aos seguintes dados:

- a) p_0 variando de 0,01 a 0,4;
- b) $n = 25, 50, 100$ e 200 ;
- c) $ARL_0 = 25, 50, 100, 200, 300, 500$;

⁸ Função uniroot() do R, baseada em 'zeroin.c', in <http://www.netlib.org/c/brent.shar> (BRENT, 1973).

d) s (alteração em p_0) variando de 1 a 100%.

Calculou-se o erro relativo das aproximações 5.15 e 5.16 em relação ao resultado de 5.17. Foram anotados os valores que correspondem ao s mínimo em que os erros relativos (%) entre os resultados das Equações e valores para h com um método iterativo são inferiores a 2% e 5% respectivamente.

Verificou-se que a acurácia destas Equações têm relação proporcional à quantidade np_0ARL_0 (5.18). Quanto mais próxima de zero é esta quantidade, menos acurada será a aproximação. O Exemplo 5.1 ilustra o exposto no neste parágrafo.

$$qc = np_0ARL_0. \quad (5.18)$$

Exemplo 5.1: Seja $p_0 = 0,01$, o tamanho da amostra $n = 50$, $p_1 = 0,011$ ($i = 10\%$) e $ARL_0 = 100$. O valor de qc é igual a 50. O limite superior h calculado pela Equação (5.15) é 4,7 e, extraindo a raiz é 5,8. O erro relativo percentual é 19%. Se o valor do ARL_0 for igual a 200, $qc = 100$; $h = 7,7$ (Eq.5.15) e $h = 8,1$ (extração da raiz). O erro relativo é de 4,9%. Se o valor do ARL_0 for igual a 300, $qc = 150$, $h = 9,6$ (Eq.5.15) e $h = 9,8$ (extração da raiz). O erro relativo fica em torno de 2%.

Como é importante definir as situações em que as aproximações são aplicáveis, verificou-se a relação de (5.15) em termos do acréscimo (s %) em p_0 . Foram identificadas combinações de valores mínimos de p_0 , n , ARL_0 e s (%) para atender um erro relativo pré-especificado.

Na Tabela 5.1 estão alguns valores de qc mínimos recomendáveis e respectivos acréscimos (s %) em p_0 para aproximações com erro relativo inferior a 2% e 5%, respectivamente. Assim, para um erro de 2%, se $qc = 125$, p_1 deve ser no mínimo $1,10p_0$ ($s \geq 10\%$). Para $qc = 600$, p_1 deve ser no mínimo $1,05p_0$ ($s \geq 5\%$). Se $qc \geq 12000$ a Equação 5.15 pode ser aplicada para quaisquer valores de s (%).

Os valores da Tabela 5.1 não pretendem ser “linhas de corte rígidas”. Esta escolha depende da precisão desejada para os parâmetros do gráfico. Se, por exemplo, $qc = 145$, isto significa que o erro relativo não deve ser muito maior que 2% para o mesmo valor de s . Para um s um pouco maior, o erro é aproximadamente 2%. O Exemplo 5.2 explana essa situação.

Tabela 5. 1–Valores mínimos de qc para considerando s (acrécimo em p_0) e erro relativo de 5% e 2%.

qc		s (acrécimo em p_0)
Erro relativo < 5%	Erro relativo < 2%	
100	150	10%
450	600	5%
1400	1500	3%
9000	12000	1%

Exemplo 5.2: Seja $p_0 = 0,01$, o tamanho da amostra $n = 50$, $p_1 = 0,011$ ($s = 10\%$) e $ARL_0 = 260$. O valor de qc é igual a 130. O limite superior h calculado pela Equação (5.15) é 8,9 e, extraindo a raiz é 9,2. O erro relativo é 3,2%. Para $s = 12\%$, $p_1 = 0,012$, nas mesmas condições, o erro será de 1,25%.

Com base nos valores da Tabela 5.1 verificou-se como ficariam alguns cenários, sendo fixado o tamanho de amostra n , ARL_0 e p_0 . Para $ARL_0 = 370$, comum em gráficos de controle, a Equação 5.16 pode ser aplicada, com erro inferior a 2%, para $s_{min} = 5\%$ ($p_1 \geq 1,05p_0$), $n \geq 50$, $ep_0 \geq 0,01$. Para $ARL_0 = 200$, a Equação 5.15 pode ser aplicada, com erro inferior a 5%, para $s_{min} = 5\%$ ($p_1 \geq 1,05p_0$), $n \geq 50$, $ep_0 \geq 0,01$. Para um erro inferior a 2% o tamanho da amostra n deve ser superior a 100. Para $ARL_0 = 100$, a Equação 5.15 pode ser aplicada, com erro inferior a 5%, para $s_{min} = 5\%$ ($p_1 \geq 1,05p_0$), $n \geq 100$, $ep_0 \geq 0,01$. Para um erro inferior a 2%, o tamanho da amostra n deve ser superior a 200.

Como regra geral, pode-se dizer que quanto maior for o valor de qc , menores são as restrições para a aplicação da aproximação, para resultados mais acurados. Valores de qc (5.18) superiores a 600 permitem trabalhar com relativa folga para $s \geq 5\%$. Nesta faixa se encontram a maioria dos casos de aplicações práticas. Em contrapartida, recomenda-se cautela em aplicar aproximação se o valor de s for inferior a 5%. Assim, para algumas situações específicas, como p_0 , n e/ou ARL_0 muito pequenos e $s < 5\%$ recomenda-se encontrar h mediante a extração da raiz, ou utilizar outro método.

Portanto, a aproximação não é apropriada para alterações muito pequenas em p_0 . Estes resultados dependem também da magnitude de p_0 , n e ARL_0 . Valores de p_0 pequenos não irão produzir resultados acurados para valores menores de s . Para o mesmo valor de p_0 ,

aumentando ARL_0 e/ou n (tamanho da amostra) há uma sensível melhora nos resultados de h para s menores.

Para uma melhor interpretação dos resultados construiu-se um gráfico comparativo que pode ser visto na Figura 5.1, para $p_0 = 0,05, n = 100$ e $ARL_0 = 300$. A linha em preto corresponde à aproximação calculando a raiz da Equação (5.17) por um método numérico. Quanto maior o valor de p_1 mais acurada é a aproximação, inclusive com menos termos. Para valores maiores de ARL_0 e/ou n (tamanho da amostra), há uma sensível melhora nos resultados de h para s menores.

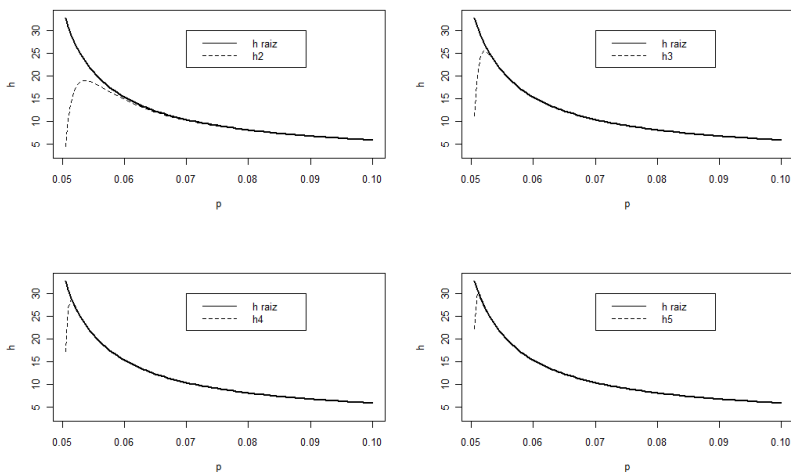


Figura 5.1— Comparação entre valores de h : raiz (h raiz) e das aproximações h_2 (2 termos), h_3 (3 termos, - Equação 5.12), h_4 (4 termos - Equação 5.15) e h_5 (5 termos-Equação 5.16).

Comparando os gráficos (Figura 5.1) verifica-se que as Equações (5.15) e (5.16) são mais acuradas do que da Equação 5.12, para um mesmo valor de $p = p_1$, ou $s(\%)$, o que é esperado em virtude do maior número de termos. Já a Equação (5.15), com um termo a menos que a Equação (5.16) tem resultados similares, com poucas variações. Os resultados com apenas dois termos (h_2) (Equação 5.10), são menos acurados para uma faixa maior de valores que as demais aproximações.

Na próxima seção os resultados das aproximações serão comparados com valores da literatura.

5.3 COMPARAÇÃO COM VALORES DA LITERATURA

Esta seção tem por objetivo analisar os resultados da aplicação das Equações (5.12), (5.15) e (5.16) em exemplos da literatura. Na Tabela 5.2 visualizam-se os resultados para o cálculo de h utilizando-se as Equações citadas aos dados de Gan (1993). O valor de p_0 (sob controle) é igual a 0,05 e para p_1 (fora de controle) assumem-se os valores 0,053; 0,056; 0,059; 0,062; 0,065 e 0,068 (primeira coluna). O tamanho da amostra é $n = 100$ e o ARL_0 é 240. Gan (1993) aproximou os parâmetros usando a abordagem de Markov (BROOK e EVANS, 1972). O valor de referência k^9 foi calculado com a Equação 3.11. O limite superior h (GAN, 1993) está na terceira coluna. Nas demais colunas estão os valores aproximados de h pelas Equações (5.12), (5.15) e (5.16) e respectivos erros relativos percentuais.

Tabela 5. 2 - Valores de h aproximados para comparação

p_1	k	h	Eq. 5.12	Er(%)	Eq. 5.15	Er(%)	Eq. 5.16	Er(%)
0,053	5,15	23,75	22,06	7,12	22,83	3,87	22,66	4,59
0,056	5,29	18,29	17,99	1,64	17,97	1,75	17,95	1,86
0,059	5,44	15,00	15,04	0,27	15,02	0,13	15,02	0,13
0,062	5,58	13,14	13,00	1,07	12,99	1,14	12,99	1,14
0,065	5,72	11,57	11,51	0,52	11,50	0,61	11,50	0,61
0,066	5,77	11,00	11,09	0,82	11,09	0,82	11,09	0,82
0,068	5,86	10,25	10,36	1,07	10,36	1,07	10,36	1,07

Na Figura 5.3 pode ser visualizado um gráfico comparativo dos valores da Tabela 5.2. Nota-se que os valores se distanciam muito pouco dos de h , do artigo de Gan (1993). Comparando-se com os valores aproximados com os de h (GAN, 1993), o erro relativo é inferior a 2%, para todos os resultados, exceto para $p_1 = 0,053$. Neste caso, a Equação (5.15) tem menor erro relativo (3,87%). Para a Equação (5.16) o erro é de 4,59% e na Equação (5.12) o erro é de aproximadamente 7,12%.

⁹Há pequenas diferenças entre alguns valores de k de Gan (1993) e calculados pela fórmula do SPRT, na segunda casa decimal. Optou-se pela fórmula por ser recomendada por Gan (1993) e Hawkins e Olwell (1998)

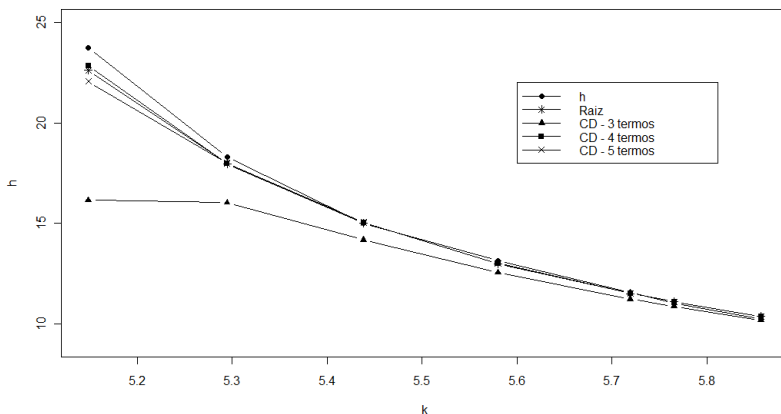


Figura 5. 2– Gráfico comparativo dos valores de h (Tabela 5.2).

Um segundo exemplo para comparação remete ao trabalho de Hawkins (1992). Na época não havia literatura que permitisse confirmar valores de ARL_0 's para CUSUM com distribuição binomial. Hawkins (1992) compara seus resultados com aproximações à distribuição de Poisson. Considerou $p_0 = 0,0024$; $n = 1000$; $k = 3$ ($p_1 = 0,0037$; $s = 54\%$) e $h = 10$. O ARL_0 calculado com o algoritmo de Hawkins é 554, foi considerado consistente pelo autor comparado com o valor assintótico de 548. Nas três aproximações o valor de h é 9,89. Recalculando o ARL_0 com o algoritmo de Hawkins e $h = 9,89$ o valor de ARL_0 também é 554. Hawkins (1992), ainda, calcula o $ARL_0=871$, para $p_0 = 0,096$; $n = 25$; $k = 3$ ($p_1 = 0,1472$; $s = 53\%$) e $h = 1$, mas não efetua comparações. Aproximando tais valores com as Equações (5.12), (5.15) e (5.16), h resulta em 9,87 (5.12) e 9,86 (5.15 e 5.16) respectivamente. Recalculando, o ARL_0 é 873 e 886 (5.15 e 5.16).

Em seguida comparam-se os resultados das aproximações com o exemplo adaptado de Morais e Pacheco (2006). Considera-se $p_0 = 0,02$, $n = 100$, $k = 3$ ($p_1 = 0,0427685$; $s = 114\%$) e $h = 6$. O ARL_0 calculado com o algoritmo de Hawkins (1992) é 459,4. Avaliando h pelas aproximações tem-se $h = 6$ (5.12), (5.15) e (5.16).

Pode-se concluir que para os exemplos da literatura as aproximações geraram bons resultados. Todavia, é importante avaliar as aproximações num espectro maior: valores diferentes de p_0 , p_1 , n e ARL_0 . E, também, se para um determinado parâmetro h calculado, as

aproximações retornam o resultado, ARL_0 esperado. Isto será abordado na seção seguinte.

5.4 COMPARAÇÃO COM VALORES SIMULADOS

Para verificar se as fórmulas são acuradas e precisas seu resultado deve ser similar ao esperado reproduzido quando aplicado a uma situação real ou simulada que corresponde às condições de projeto. Foram então realizadas duas análises. Uma faz comparação com os valores produzidos pelo algoritmo de Hawkins¹⁰, e a outra utiliza simulação. Na Tabela 5.3 encontram-se os resultados para um ARL_0 alvo de 300, $p_0 = 0,05$ e $n = 100$. Utilizou-se apenas a Equação (5.15) nesta análise. A Equação (5.15) tem menos termos que a Equação (5.16), portanto menos operações a serem feitas e, nas análises anteriores, tem resultados próximos. Na simulação foram geradas 3000 amostras com distribuição binomial de tamanho $n = 100$ e probabilidade p_0 . Aplicou-se um gráfico CUSUM com o parâmetro k calculado pela Equação (3.11) e o limite superior h pela Equação (5.15). Anotou-se o número da amostra onde há a primeira indicação de um alarme falso. Este procedimento foi replicado $r = 20000$ vezes. Foi calculada a média dos r valores anotados (ARL_{S_0}). Um fluxograma com os passos da simulação se encontra na Figura 5.4. A rotina se encontra no Apêndice C.

Percebe-se, mediante os valores da Tabela 5.3, que a aproximação não apresenta bons resultados para grandes alterações ($s \geq 70\%$) no cenário analisado. O erro relativo percentual, calculado em relação ao ARL_0 alvo de 300, aumenta na medida em que s é maior. Este comportamento tem explicação na literatura.

As fórmulas têm origem nos trabalhos de Siegmund (1985). A aproximação de Siegmund para o ARL é conhecida por ser pouco acurada para valores maiores de k (HAWKINS e OLWELL, 1998; ROGERSON, 2006). Os valores simulados ficam próximos dos resultados da abordagem markoviana, para toda a faixa analisada.

¹⁰Foi aplicada a função `arICUSUM` do pacote `SURVEILLANCE` (HÖHLE, 2007) no GNU R, baseada no código em FORTRAN de Hawkins (1992).

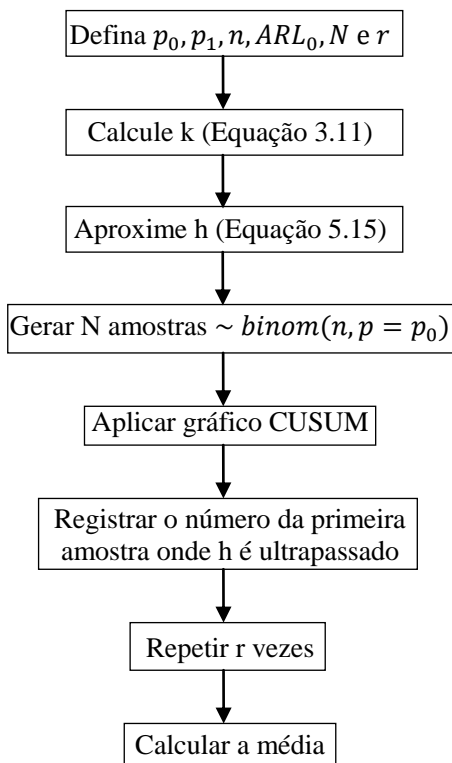


Figura 5. 3– Fluxograma do processo de simulação para ARL.

Tabela 5. 3 - Valores de ARL_0 com algoritmo de Hawkins (ARL_{c_0}) e simulados (ARL_{s_0}) para h (aproximado com a Equação 5.15) e k (calculado pela Equação 3.11) para um ARL_0 alvo de 300).

p_1	s (%)	k	h	ARL_{c_0}	ARL_{s_0}	er (%)
0,055	10	5,25	20,9	296	293	2,33
0,06	20	5,49	15,3	301	297	1,00
0,065	30	5,72	12,2	287	286	4,67
0,07	40	5,95	10,3	294	292	2,67
0,075	50	6,17	9,0	312	308	2,67
0,08	60	6,39	8,0	310	311	3,67
0,085	70	6,61	7,3	341	339	13,00
0,09	80	6,82	6,7	369	366	22,00
0,095	90	7,03	6,2	458	455	51,67
0,1	100	7,24	5,8	469	478	59,33

A Tabela 5.4 traz valores de ARL_0 simulados e calculados por cadeias de Markov para os mesmos valores de p_0, p_1 e n , mas com $ARL_0 = 370$, valor comum para processos industriais. Verifica-se o mesmo comportamento observado no caso anterior: o h calculado pela aproximação CD aproxima razoavelmente os valores de ARL_0 para pequenas a moderadas alterações em p_0 . Na medida em que s (p_1) aumenta, os resultados se distanciam do alvo. Observa-se novamente que os valores resultantes da abordagem markoviana acompanham os obtidos por simulação. Foram geradas $N = 3000$ amostras de tamanho $n = 100$ com $p = p_0$ e repetiu-se o experimento $r = 20000$.

Tabela 5. 4 - Valores de ARL_0 com algoritmo de Hawkins (ARL_{C0}) e simulados (ARL_{S0}) com h (Equação 5.15) e k (Equação 3.11) para ARL_0 alvo de 370, $n = 100$ e $p_0 = 0,05$.

p_1	s	k	h	ARL_{C0}	ARL_{S0}
0,055	10	5,25	22,51	373,99	370,2
0,06	20	5,49	16,25	370,92	372,3
0,07	40	5,95	10,88	372,03	370,2
0,09	80	6,82	6,99	431,64	433,2
0,10	100	6,06	7,24	514,72	517,8

Na Tabela 5.5 pode ser visto o desempenho em termos de ARL_1 para os mesmos parâmetros. Na primeira coluna estão os valores de p_1 e na segunda coluna, a alteração (s %) em $p_0 = 0,05$. Os valores de k estão na terceira coluna e os valores de h , aproximados pela Equação 5.15, na coluna quatro. Na coluna cinco estão os valores ARL_1 (ARL_C) calculados com cadeias de Markov e na sétima coluna os valores de ARL_1 (ARL_S) obtidos por simulação. O erro relativo percentual (er) vem em seguida.

Especificamente para este caso, considerando h aproximado pela Equação (5.15), os valores obtidos por simulação do ARL_1 são muito bons, quando comparados aos obtidos por cadeias de Markov. O erro relativo percentual é inferior a 1% quando comparado com os obtidos por Hawkins (1992). Uma diferença no processo de simulação foi observada. Inicialmente, foi gerada uma sequência de $N = 30$ amostras sob controle, com $p = p_0$. Em seguida N amostras com $p = p_1$, sendo $p_1 = 0,05, N = 1000$. Para os demais casos, $N = 500$. O procedimento CUSUM foi aplicado e foi anotado o número da amostra em que a estatística calculada ultrapassa h , considerando que uma

mudança ocorre na 31ª amostra. Isto foi necessário para que fosse possível comparar os resultados com cadeias de Markov. Na abordagem markoviana considera-se o processo sob controle no início (TRINDADE, 2008).

Os resultados observados indicam que a expressão pode não ser apropriada para valores maiores de k . O valor de referência k , ao contrário de um CUSUM para monitoramento da média, não está vinculado a uma medida simples – o desvio em relação à média. Para variáveis mensuráveis, a Equação de Siegmund não é apropriada para $k \geq 1$ (HAWKINS e OLWELL, 1998) e as aproximações de Rogerson (2006) também são menos acuradas para valores de $k > 1$ e $k < 1/ARL_0$. Já para variáveis binomiais a proporção sob e fora de controle estatístico, além do tamanho n da amostra influenciam o valor de referência k , tornando mais difícil classificar o que é um valor de referência grande.

Tabela 5. 5 - Valores de ARL_1 simulados (ARL_{C1}) e calculados (ARL_{S1}) para $p_0 = 0,05$, p_1 , n e $ARL_0 = 370$.

n	p_1	s (%)	k	h	ARL_{C1}	ARL_{S1}	er(%)
50	0,055	10	2,62	18,61	91,83	91,90	0,07
50	0,060	20	2,74	14,28	43,38	43,52	0,32
50	0,065	30	2,86	11,78	26,22	26,28	0,23
50	0,070	40	2,97	10,14	17,54	17,57	0,17
50	0,075	50	3,09	8,97	13,32	13,38	0,45
100	0,055	10	5,25	22,51	63,96	63,95	0,03
100	0,060	20	5,49	16,25	27,38	27,40	0,07
100	0,065	30	5,72	12,95	15,63	15,60	0,19
100	0,070	40	5,95	10,88	10,51	10,52	0,09
100	0,075	50	6,17	9,46	7,61	7,60	0,13

Os seguintes exemplos ilustram o problema:

i) $p_0 = 0,02$; $p_1 = 0,0427$; $s = 114\%(1,6\sigma)$; $n = 100$; $ARL_0 = 459,4$ e $k = 3$. O valor de h obtido por aproximação é $h = 6$. Recalculando o ARL_0 é aproximadamente 459,36.

ii) $p_0 = 0,08$; $p_1 = 0,088$; $s = 10\%(0,21\sigma)$; $n = 50$; $ARL_0 = 200$ e $k = 4,2$. O valor aproximado de h é 16,5 e o ARL_0 recalculado é 202,5.

Nota-se que no primeiro caso o valor de k é menor ($3 < 4,2$), embora a magnitude da mudança seja maior ($114\% \gg 10\%$). Assim, aproximação da Eq. (5.15) não é recomendada para valores:

- a) de s muito pequenos, pela própria condição da aproximação (ver Seção 5.2), um valor sugerido é $s \geq 10\%$;
- b) de k maiores que 10, devendo-se ter cautela com valores de $5 < k \leq 10^{11}$;
- c) de $p_0 < 0,01$ (ver Subseção 3.8.4).

5.5 USO DA APROXIMAÇÃO CD COM INTERPOLAÇÃO

Nas seções anteriores verificou-se que as aproximações desenvolvidas (em particular a da Equação 5.15) não são acuradas para valores grandes de k e devem obedecer a certas condições (Equação 5.18). Nesta seção é apresentado um algoritmo que, mediante um método de interpolação amplia a faixa de utilização da Equação 5.15. Em verdade, o h aproximado servirá como um valor inicial para um processo iterativo que abrange a abordagem markoviana. É uma forma simples de contornar as deficiências anotadas.

Partiu-se do algoritmo da função “findH()”(Anexo D) do pacote surveillance (HÖHLE, 2007), que aproxima valores de h e k para um CUSUM binomial por cadeias de Markov. Nesta função é estabelecido como valor inicial, o mínimo entre doze (12) e quatro vezes o valor de referência ($4k$). Forma-se então uma Tabela com valores de h próximos ao valor inicial e os respectivos valores de ARL, calculados por cadeias de Markov (HAWKINS, 1998). Através de interpolação logarítmica é aproximado um valor de h para o ARL_0 especificado. Inicia-se um processo iterativo, em que novos valores de h são gerados e aproximados por interpolação até que o critério de convergência (erro máximo tolerável) seja alcançado.

A modificação proposta está em considerar o h inicial aquele encontrado pela Equação (5.15), quando o erro entre o valor do ARL calculado com o parâmetro h for maior que um valor tolerável pré-especificado. Deste modo é possível diminuir a quantidade de iterações

¹¹Mais estudos são necessários aqui. Os valores de n, p_0, s e ARL_0 têm influência nos resultados.

(e o tempo de computação) no algoritmo de Hohle (2007). Além disso, as limitações da aproximação são contornadas. Os procedimentos estão dispostos a seguir:

São dados p_0, p_1, n, ARL_0 , erro máximo tolerável e número de dígitos para k e h :

1. O valor de referência k é calculado pela Expressão (3.11);
2. O limite superior h é estimado pela Equação (5.15);
3. Calcula-se o ARL_0 com os parâmetros calculados por cadeias de Markov;
4. Obtém-se o erro relativo entre o ARL_0 calculado e o ARL_0 de projeto;
5. Se o erro for inferior ao erro máximo tolerável então Limite superior do CUSUM = h ;
6. Enquanto o erro relativo for maior ou igual ao valor erro máximo tolerável:

a. $h_1 = h + 1$;

b. $h_2 = h - 1$;

c. Calcula-se o valor de ARL com h_1 e h_2 ;

d. Aproxima-se h mediante interpolação logarítmica

com os valores de h_1, h_2, ARL_{01} e ARL_{02} para $ARL = ARL_0$.

e. Calcula-se o erro relativo.

O programa retorna o limite superior h , o valor de referência k e ARL_0 . Um fluxograma dos procedimentos se encontra na Figura 5.4.

Verificou-se que na maioria dos casos que em que o h aproximado não retornava um ARL próximo ao desejado, com pequenos ajustes (arredondamento para o inteiro superior ou inferior), o valor de projeto era alcançado (com um erro máximo tolerável). Em virtude disto decidiu-se que o incremento no processo de interpolação estivesse no intervalo $h \pm 1$. Desta forma pretende-se diminuir o processo iterativo e obter melhores aproximações. Apenas para comparação, o intervalo considerado no processo de interpolação no pacote *surveillance* (Anexo D) é $h_1 = 2h$ e $h_2 = h_1 \wedge \text{sign}(ARL_0 - ARL_c)$, onde ARL_0 é ARL de projeto, ARL_c o calculado e sign corresponde ao sinal obtido. Foi elaborado um programa no R (hrs.5) e, nas Tabelas 5.7, 5.8 e 5.9 estão dispostos alguns resultados do uso do algoritmo. As Tabelas trazem: os valores de ARL_0 alvo, k e h estimados pela função *hrs.5*, ARL_0 determinado pelo algoritmo de Hawkins (ARL_{c0}) e N número de amostras utilizadas na simulação. A rotina se encontra no Apêndice D.

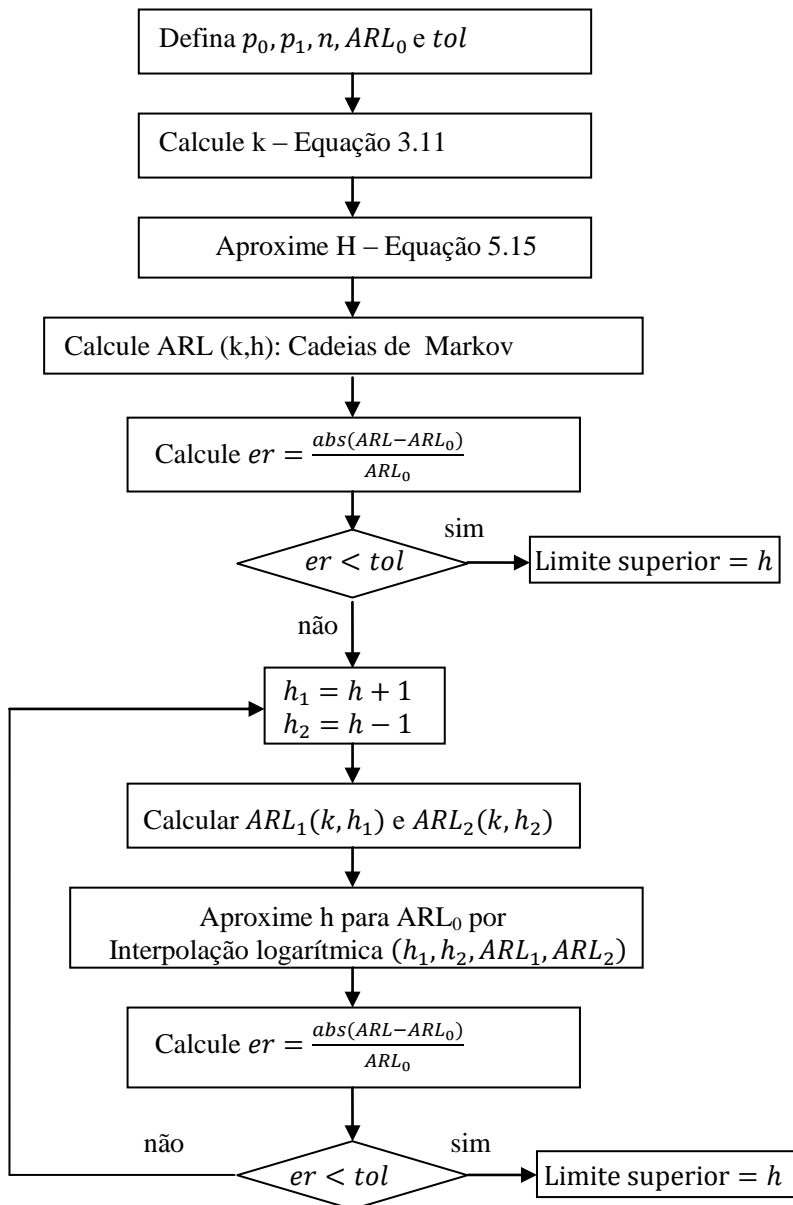


Figura 5. 4 – Fluxograma do processo de aproximação do limite superior do gráfico CUSUM binomial.

Os valores obtidos com o programa hrs.5 foram comparados com os resultados da abordagem markoviana. De modo geral, uma ou duas iterações são necessárias na maioria dos casos em que o erro é maior que o tolerável. Apenas para valores de k muito grandes ($k = 86$, por exemplo) são realizadas mais iterações. Esse processo também depende do número de casas decimais de k e h . Quanto mais decimais, mais lento é o processo, em virtude da parte que usa cadeias de Markov, quando necessária.

A Tabela 5.6 traz os valores de h aproximados pelo programa hrs.5 para alguns valores de p_0 sob controle $0,01 \leq p_0 \leq 0,1$, $s = 10\%$, ARL_0 variado e tamanho de amostra $n = 100$ e 200 .

Tabela 5. 6 - Valores de h calculados pelo programa hrs.5, $ARL_0(ARLc_0)$ calculado por cadeias de Markov e erro relativo (%), para $0 \leq p_0 \leq 0,1$; $s = 10\%$, com ARL_0 alvo variado e tamanho de amostra $n = 100$ e 200 .

n	ARL_0 alvo	p_0	k	h	$ARLc_0$	erro (%)
100	50	0,01	1,05	5,3	51,5	2,91
100	100	0,01	1,05	7,5	100,71	0,70
100	200	0,01	1,05	10,4	200,55	0,27
100	300	0,01	1,05	12,5	303,85	1,27
100	370	0,01	1,05	13,6	370,47	0,13
100	500	0,01	1,05	15,4	502,15	0,43
200	50	0,01	2,1	7,0	49,84	0,32
200	100	0,01	2,1	10,0	100,72	0,71
200	200	0,01	2,1	13,6	199,7	0,15
200	300	0,01	2,1	16,1	300,53	0,18
200	370	0,01	2,1	17,5	371,6	0,43
200	500	0,01	2,1	19,6	502,26	0,45
100	50	0,025	2,62	7,7	50,5	0,99
100	100	0,025	2,62	10,9	101,07	1,06
100	200	0,025	2,62	14,8	201,43	0,71
100	300	0,025	2,62	17,4	300,16	0,05
100	370	0,025	2,62	18,9	372,3	0,62
100	500	0,025	2,62	21,1	501,43	0,29
200	50	0,025	5,25	10	49,24	1,54
200	100	0,025	5,25	13,8	101,34	1,32
200	200	0,025	5,25	18,4	200,31	0,15
200	300	0,025	5,25	21,3	300,99	0,33

Tabela 5.6 - Continuação.

200	370	0,025	5,25	23,1	377,02	1,86
200	500	0,025	5,25	25,4	498,32	0,34
200	50	0,05	10,49	12,5	48,17	3,80
200	100	0,05	10,49	17,1	99,73	0,27
200	200	0,05	10,49	22,2	199,3	0,35
200	300	0,05	10,49	25,5	296,47	1,19
200	370	0,05	10,49	27,3	370,3	0,08
200	500	0,05	10,49	29,9	499,88	0,02
100	100	0,075	7,87	15,3	99,78	0,22
100	200	0,075	7,87	20,1	200,9	0,45
100	300	0,075	7,87	23,2	301,39	0,46
100	370	0,075	7,87	24,8	368,99	0,27
100	500	0,075	7,87	27,3	501,91	0,38
200	50	0,075	15,74	13,9	49,54	0,93
200	100	0,075	15,74	16,4	98,98	1,03
200	200	0,075	15,74	24,1	198,68	0,66
200	300	0,075	15,74	27,5	299,64	0,12
200	370	0,075	15,74	29,2	366,2	1,04
200	500	0,075	15,74	31,9	498,44	0,31
100	50	0,1	10,49	12,3	52,05	3,94
100	100	0,1	10,49	16,3	98,87	1,14
100	200	0,1	10,49	21,4	200,84	0,42
100	300	0,1	10,49	24,6	300,36	0,12
100	370	0,1	10,49	26,2	369,75	0,07
100	500	0,1	10,49	28,7	500,8	0,16
200	50	0,1	20,99	15,2	55,42	9,78
200	100	0,1	20,99	19,7	97,72	2,33
200	200	0,1	20,99	25,0	186,85	7,04
200	300	0,1	20,99	28,6	304,71	1,55
200	370	0,1	20,99	30,1	386,5	4,27
200	500	0,1	20,99	32,8	488,4	2,38

A Tabela 5.7 traz os valores calculados pelo programa hrs.5 para valores de $p_0 \geq 0,1$, $s = 15\%$, ARL_0 variável e tamanho de amostra $n = 100$ e 200 .

A Tabela 5.8 traz os valores calculados pelo programa hrs.5 para valores de $p_0 \leq 0,01$ (processos de alto rendimento), $s = 25\%$, tamanho de amostra $n = 100$ e 200 .

Tabela 5.7 - Valores de h calculados pelo programa hrs.5, $ARL_0(ARLc_0)$ calculado por Markov e erro relativo (%), com $p_0 > 0,1$; $s = 15\%$, ARL_0 alvo variado e tamanho de amostra $n = 100$ e 200 .

n	p_0	ARL ₀ alvo	k	h	ARLc ₀	Erro (%)
100	0,2	50	21,47	11,3	52,10	4,03
100	0,2	200	21,47	18,1	198,34	0,84
100	0,2	370	21,47	21,3	369,21	0,21
200	0,2	50	42,95	12,4	51,11	2,17
200	0,2	200	42,95	19,8	201,7	0,84
200	0,2	370	42,95	23,2	366,69	0,90
100	0,3	50	32,22	10,6	50,85	1,67
100	0,3	200	32,22	16,8	198,76	0,62
100	0,3	370	32,22	19,6	365,09	1,34
200	0,3	50	64,45	11,5	48,43	3,24
200	0,3	200	64,45	17,8	202,25	1,11
200	0,3	370	64,45	20,75	371,17	0,32
100	0,4	50	42,98	9,2	50,98	1,92
100	0,4	200	42,98	14,7	190,30	5,10
100	0,4	370	42,98	17,6	398,80	4,70
200	0,4	50	85,97	9,6	47,23	5,86
200	0,4	200	85,97	15,1	205,99	5,99
200	0,4	370	85,97	17,6	359,62	2,89

Como pode ser verificado nas Tabelas 5.6, 5.7 e 5.8, o algoritmo desenvolvido é mais uma alternativa à aproximação do limite superior do CUSUM binomial. Os valores de erros relativos são aceitáveis, sendo na maioria dos cálculos realizados inferior a 5%. Erros relativos maiores estão associados, em geral, a valores maiores de k ($k > 20$, por exemplo). Alguns pontos positivos podem ser verificados:

- a) A aproximação auxilia a encontrar o limite superior h diretamente em alguns casos;
- b) este valor aproximado é um bom valor inicial para um processo iterativo para calcular h .

A abordagem markoviana é excelente para o cálculo do ARL, quando h é conhecido. Embora não fosse um objetivo deste trabalho, pode ser visto, a partir dos resultados das simulações o quão próximos são os resultados. O caminho inverso (determinar h , para um ARL conhecido), no entanto, pode ser muito complexo. Para exemplificar pode-se remeter à função findH do pacote surveillance do R, já citado. O

programa “Anyarl” de Hawkins (1998) permite encontrar valores de h , sendo dado o ARL e k . Mas nem todos os valores de k são possíveis de utilização, sendo necessário ajustá-los, o que implica alterar a proporção p_1 de planejamento. A aproximação CD de Reynolds e Stoumbos (1999) quando coloca k e h como múltiplos de um valor l , também implica o ajuste de p_1 , para que isto seja possível. O algoritmo desenvolvido permite que se mantenha a proporção p_1 de projeto.

Tabela 5. 8 - Valores de h calculados pelo programa hrs.5, ARL_0 calculado por cadeias de Markov e erro relativo (%), para $p_0 \leq 0,01$; $s = 25\%$, ARL_0 variado e tamanho de amostra $n = 100$ e 200 .

n	ARL_0 a	p_0	k	h	ARL_0 c	Erro (%)
100	50	0,001	0,11	1,7	50,13	0,26
100	100	0,001	0,11	2,6	108,23	8,23
100	200	0,001	0,11	3,7	213,16	6,58
100	370	0,001	0,11	4,9	384,20	3,84
200	50	0,001	0,22	2,5	53,92	7,84
200	100	0,001	0,22	3,6	106,80	6,80
200	200	0,001	0,22	5,0	210,65	5,33
200	370	0,001	0,22	6,5	380,26	2,77
100	50	0,0025	0,28	2,7	53,07	6,14
100	100	0,0025	0,28	3,8	102,20	2,20
100	200	0,0025	0,28	5,2	200,37	0,19
100	370	0,0025	0,28	6,8	379,88	2,67
200	50	0,0025	0,56	3,6	51,70	3,40
200	100	0,0025	0,56	5,0	102,69	2,69
200	200	0,0025	0,56	6,7	201,22	0,61
200	370	0,0025	0,56	8,5	369,87	0,04
100	50	0,005	0,56	3,6	51,83	3,66
100	100	0,005	0,56	5,0	103,02	3,02
100	200	0,005	0,56	6,7	202,03	1,02
100	370	0,005	0,56	8,5	372,61	0,71
200	50	0,005	1,12	4,5	48,65	2,70
200	100	0,005	1,12	6,3	101,66	1,66
200	200	0,005	1,12	8,3	198,60	0,70
200	370	0,005	1,12	10,4	368,36	0,44

É importante ressaltar que o valor exato do ARL_0 desejado nem sempre consegue ser atingido, face às próprias características da distribuição binomial e do comportamento do CUSUM.

5.6 CONSIDERAÇÕES FINAIS

Embora não fosse um objetivo deste trabalho, verificou-se nas simulações que a abordagem por cadeias de Markov produz bons resultados para o cálculo do ARL. No começo dos trabalhos pairavam ainda dúvidas sobre com o que comparar os resultados. A literatura inclusive reforçava isso. A partir de certo momento, pode-se optar pela comparação unicamente com os resultados produzidos pela abordagem markoviana.

5.7 SÍNTESE DO CAPÍTULO

Neste capítulo foram desenvolvidas fórmulas diretas para o cálculo de h com a aproximação CD de Reynolds e Stoumbos (1999, 2001) utilizando a função LambertW. Uma expressão, com 4 termos, produziu valores acurados quando comparada ao valor de h encontrado métodos numéricos, para valores de s não muito pequenos. A fórmula foi avaliada em termos acurácia comparando-se os resultados com a literatura e fazendo-se uso de simulação. A aplicação das fórmulas em exemplos da literatura existente também apresentou bons resultados. A simulação permitiu identificar que a aproximação não responde bem para grandes valores de k . Todavia não foi possível, neste momento, associar a acurácia da aproximação tendo como base unicamente a magnitude de mudança desejada, pois são muitas as especificidades. Assim, como procedimento alternativo, um h aproximado pela Equação (5.15) pode servir de valor inicial para um processo iterativo para o cálculo do limite superior h do CUSUM. Elaborou-se então um algoritmo que em que o valor de h estimado fosse refinado mediante interpolação logarítmica e cadeias de Markov para o cálculo do ARL. Testou-se o procedimento em diversos casos e aproximações para h retornaram valores de ARL_0 muito próximos dos projetados.

6. DESEMPENHO DE UM GRÁFICO COMBINADO SHEWHART-CUSUM BINOMIAL

Neste capítulo é estudada a distribuição do RL (*run length*) de um gráfico combinado e o seu desempenho comparado com o gráfico CUSUM, para os mesmos parâmetros, e em alguns casos com o gráfico Shewhart. Pretende-se analisar a forma da distribuição do RL, ou seja, se os parâmetros de projeto são atingidos e principalmente se há aumento na sensibilidade do CUSUM para alterações maiores do que as de planejamento do CUSUM, com a adição do limite superior de Shewhart. Procedimentos de simulação são aplicados para atingir estes objetivos. Para efeitos deste trabalho a parte CUSUM do gráfico combinado será planejada considerando-se uma proporção p_1 de pequena magnitude, com valores de s inferiores a 50% e em termos de desvio padrão, inferiores a $1,0\sigma$.

6.1 DISTRIBUIÇÃO DO RL DO GRÁFICO COMBINADO

Nesta seção investiga-se a distribuição do RL do gráfico combinado. Deseja-se verificar, para os parâmetros do gráfico, se o ARL_0 realmente é o de projeto, e como irá se comportar o RL (e momentos) para diversos valores de $p \neq p_0$. Para tal finalidade é analisado o comportamento do RL das partes individuais (CUSUM e Shewhart) em conjunto com o RL do gráfico combinado. Isto foi considerado importante em virtude da própria regra de decisão deste gráfico: o primeiro ponto que ultrapassa os limites de controle, ou no gráfico do tipo Shewhart, ou no gráfico CUSUM.

Os parâmetros dos gráficos são calculados de acordo com o exposto nos Capítulos anteriores: o ARL_0 do combinado e das partes individuais mediante a Equação (4.5); o valor de referência k com a Equação (3.11); o limite de decisão com a Equação (5.15) (implementado na rotina hrs.5 – Apêndice E – do R) e se o limite superior de Shewhart é convencional ou probabilístico, de acordo com as possibilidades.

Além dos gráficos e medidas usuais, como histogramas e medidas descritivas, será aplicado o gráfico violin plot (ADLER, 2005). O gráfico violin plot é uma combinação do boxplot e do gráfico de

densidade kernel. Assim, é possível visualizar a forma da distribuição, as medidas de posição e a média num mesmo gráfico. A definição das classes nos histograma pode ser complicada em virtude da amplitude dos dados (principalmente na parte Shewhart). Nestes casos o violin plot é uma alternativa gráfica para visualizar a distribuição. Mais informações sobre o violin plot estão no Anexo E.

O RL e alguns momentos são obtidos mediante simulação. Ressalta-se que com as simulações feitas, não se têm o objetivo de definir parâmetros para os gráficos e sim verificá-los. Além disso, a análise dos resultados destas simulações irá contribuir para o entendimento de como se comporta o RL do combinado. Alguns cenários de gráficos combinados foram simulados, todos de modo geral, com resultados similares. Portanto, somente dois cenários casos são mostrados nas seções seguintes. Os cenários foram escolhidos procurando-se observar valores de ARL_0 , p_0 e tamanhos de amostra variados.

6.1.1 RL do gráfico combinado - Exemplo 6.1

Sendo dados $p_0 = 0,06$; $n = 100$ e $p_1 = 0,066 = 1,1p_0$ ($s = 10\%$; $0,25\sigma^{12}$) e $ARL_0 = 370$, é projetado um gráfico combinado binomial para as especificações anteriores. Os parâmetros do gráfico combinado são: $k = 6,3$; $h = 24$ e $LSC = 15$. O ARL_0 da parte é CUSUM é 390. O limite convencional de Shewhart foi utilizado, com $LSC = 15$, ($q = 3,8$ aproximadamente) e $ARL_0 \cong 7196$ ($\alpha = 0,014\%$). Aqui a regra 1 (Seção 2.4) é atendida com $np = 6 > 5$.

Simulou-se N^{13} amostras com distribuição binomial, tamanho $n = 100$ e probabilidade $p = 0,06(p_0)$; $0,066(p_1)$; $0,072$; $0,078$; $0,084$; $0,090$; $0,096$; $0,102$; $0,108$; $0,114$; $0,120$; $0,132$; $0,150$; $0,180$; $0,210$ e $0,240$, com s variando de 0 a 300%. A estas amostras foi aplicado um gráfico combinado, registrando-se:

- i. O número da primeira amostra em que LSC é ultrapassado;
- ii. O número da primeira amostra em que h é ultrapassado;
- iii. O menor dos dois anteriores (regra de decisão do combinado).

¹²Inclui-se também, para melhor compreensão dos exemplos, a magnitude de mudança de planejamento da parte CUSUM em números de desvios padrão em relação à média (np_0).

¹³O valor de N é variado, sendo maior para p igual ou próximo a p_0 e menor para valores de p maiores. Os valores estão no Apêndice E.

Esse procedimento é replicado $r = 20000$ vezes e então estimados os valores de ARL, quartil inferior (q_1), mediana (q_2 ou MRL), quartil superior (q_3), SDRL e os coeficientes de assimetria (CSRL) e curtose (CKRL). A rotina aplicada nas simulações, desenvolvida no GNU R, está no Apêndice E. As simulações consideram o processo já iniciando fora de controle estatístico, quando $p \neq p_0$, conforme exposto na Seção 3.7.

6.1.2 Resultados – Exemplo 6.1

Na Tabela 6.1 estão os valores de ARL, CSRL e CKRL para o gráfico combinado, além dos valores de “ARL” das partes CUSUM e Shewhart em separado. Como a taxa de alarmes falsos para a parte Shewhart é muito baixa (ARL_0 alto) não foi possível alcançar este valor na simulação. Todavia o ARL_0 de projeto do gráfico combinado (368,9) ficou muito próximo ao ARL_0 de projeto (370).

Quando $p = p_0$ o RL do combinado tem praticamente a mesma forma de distribuição do RL de um gráfico CUSUM individual. É razoável, pois, o ARL_0 do gráfico combinado está muito próximo do ARL_0 do CUSUM individual. O limite h é mais estreito e o limite LSC mais relaxado. Os histogramas (Figura 6.1) e o gráfico violin plot (Figura 6.2) confirmam isto. A distribuição é assimétrica à direita (coeficiente de assimetria CSRL = 2,11) e leptocúrtica (coeficiente de curtose CKRL = 10,4). Estas características, forma assimétrica e excesso de curtose para a situação sob controle, também foram verificadas por Morais (2002), para o único caso estudado ($p_0 = 0,02$, $p_1 = 0,0427685$, $n = 100$, $k = 3$, $h = 6$ e $LSC = 4$), com CSRL=2 e CKRL = 6. Uma diferença pode ser observada em relação ao trabalho de Morais (2002) que planejou um combinado (e CUSUM) para uma alteração de aproximadamente 115% em p_0 (aproximadamente $1,6\sigma$), enquanto que nesta análise a parte CUSUM foi planejada para magnitudes de mudança menores.

O histograma e o violin plot da parte Shewhart são omitidos para este caso, em virtude da amplitude dos valores ($q_1 = 735$) e da dificuldade de se alcançar o ARL_0 de planejamento na simulação realizada. Os alarmes falsos emitidos no combinado são quase que exclusivos da parte CUSUM, em que o ARL_0 de projeto ($ARL_0 = 390$) é bem menor que da parte Shewhart ($ARL_0 = 7196$).

Tabela 6. 1- ARL, CSRL e CKRL para o gráfico combinado do Exemplo 6.1 e valores de ARL das partes CUSUM e Shewhart.

s	p (%)	Parte		Gráfico		
		CUSUM	Shewhart	Combinado	CSRL	CKRL
0	0,060	400,3	2911 ¹⁴	368,90	2,11	10,37
10	0,066	59,53	231,1	58,38	1,81	8,02
20	0,072	27,16	207,70	26,70	1,27	5,61
30	0,078	17,78	163,40	17,40	0,94	4,94
40	0,084	13,30	103,70	12,86	0,53	4,26
50	0,090	10,72	59,06	10,18	0,18	3,82
60	0,096	9,06	34,73	8,41	-0,07	3,52
70	0,102	7,90	21,83	7,08	-0,22	3,11
80	0,108	7,04	14,24	6,01	-0,24	2,64
90	0,114	6,38	9,79	5,15	-0,15	2,28
100	0,120	5,84	7,01	4,40	0,00	2,03
120	0,132	5,08	4,15	3,33	0,36	1,97
150	0,150	4,32	2,33	2,21	1,07	3,25
200	0,180	3,58	1,35	1,35	2,28	8,93
250	0,210	3,15	1,09	1,09	3,68	17,76
300	0,240	2,96	1,02	1,02	7,21	54,49

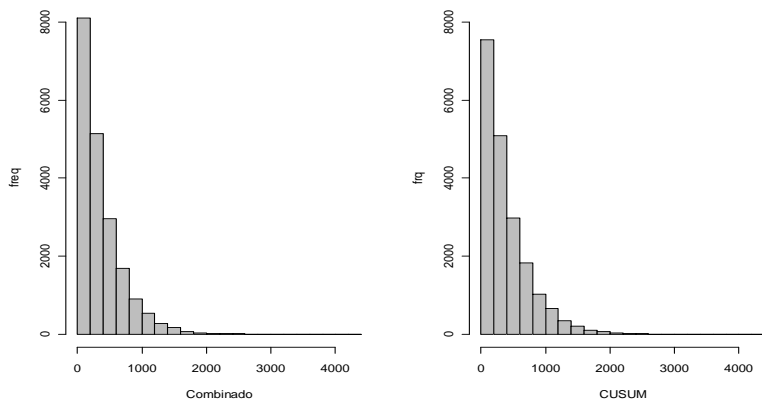


Figura 6. 1– Histogramas do valor de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 para $p = p_0(0,06)$.

¹⁴ O valor de projeto não foi alcançado pelas simulações, não sendo ultrapassado LSC em cerca de 22% das rodadas.

No gráfico violin plot (Figura 6.2) as semelhanças entre o combinado e a parte CUSUM, estão bem claras. A amplitude, os quartis e a forma da distribuição do RL do gráfico combinado, são similares as da parte CUSUM para $p = p_0$.

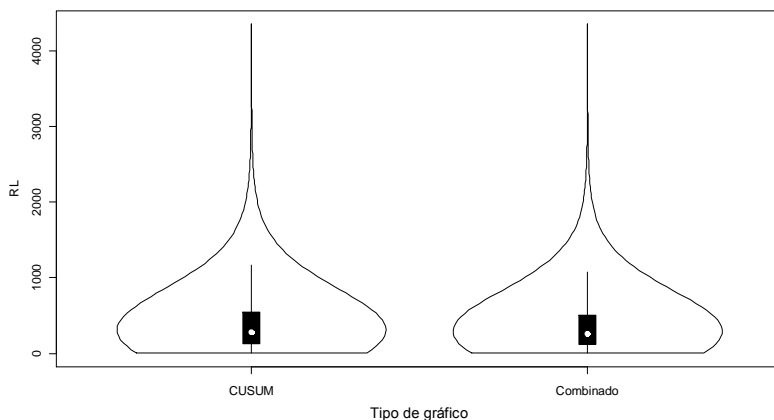


Figura 6. 2 – Gráfico violin plot dos valores de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 para $p = p_0$ (0,06).

Na análise seguinte, a situação não é mais de controle estatístico. Tem-se $p = p_1$, ou seja, a proporção fora de controle estatístico para o qual o gráfico (parte CUSUM) foi planejado para detectar. Verifica-se que a distribuição do ARL do COMBINADO é similar à da parte CUSUM (Figuras 6.3 e 6.4), o que era esperado. Como a alteração é pequena ($s = 10\%$; $0,25\sigma$) é a parte CUSUM que emite o sinal primeiro, na maior parte das vezes.

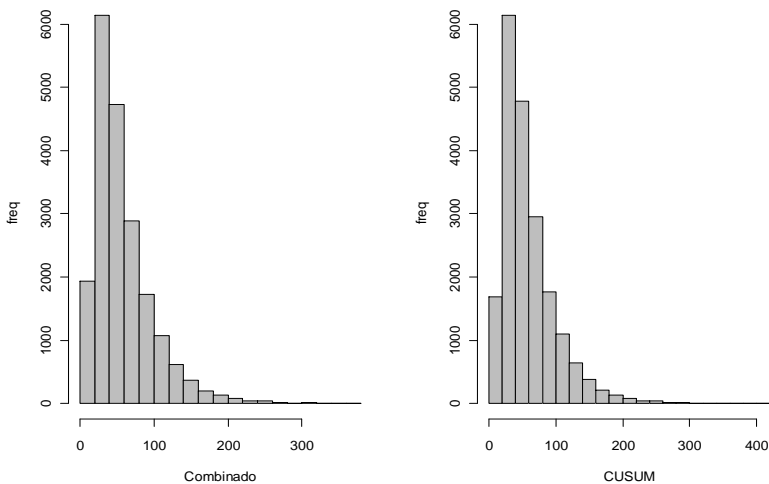


Figura 6.3 – Histogramas dos valores de ARL do procedimento combinado e da parte CUSUM para o Exemplo 6.1 com $p = p_1 = 0,066$.

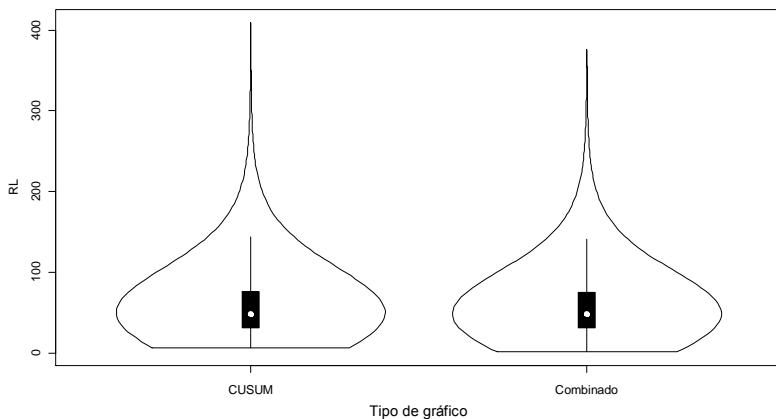


Figura 6.4 – Gráfico violin plot dos valores de ARL do gráfico combinado e da parte CUSUM para o Exemplo 6.1 com $p = p_1 = 0,066$.

Os próximos gráficos e as análises correspondentes são para valores de p maiores que p_1 . Para $p = 0,09$ ($s = 50\%; 1,26\sigma$) a distribuição do RL do combinado começa a ter características próprias.

Embora, em média a parte CUSUM detecte mais rapidamente, a forte assimetria da parte Shewhart (ver histogramas, Figuras 6.5 e 6.6) faz com que haja uma probabilidade (pequena, mas que deve ser levada em conta) de que possa sinalizar mais cedo. Isto é incorporado na distribuição do gráfico combinado, visível no gráfico violin plot (Figura 6.7). Veja que q_1 e a mediana dos dois gráficos (CUSUM e combinado) estão no mesmo nível, e q_3 é maior para o CUSUM (13) que no combinado (12). A forma distribuição do RL combinado, mais “abrangente”, é decorrente da parte Shewhart. A partir daqui percebe-se que, a parte Shewhart começa a influenciar de fato o RL. A comparação entre os histogramas (e/ou violin plots) das partes CUSUM e Shewhart com o combinado não é ainda neste ponto muito proveitosa, por causa da amplitude (grande) dos valores na parte Shewhart. O gráfico boxplot (Figura 6.8) ilustra isso.

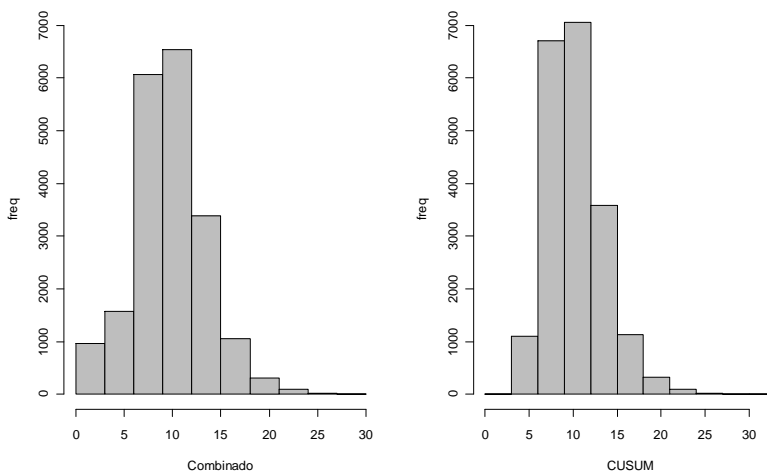


Figura 6. 5– Histogramas dos valores de ARL do gráfico combinado e somente parte CUSUM para o Exemplo 6.1 com $p = 0,09$.

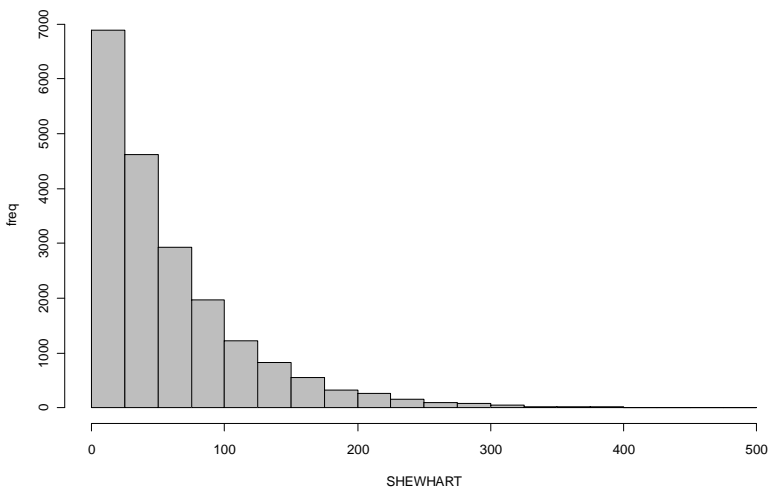


Figura 6. 6– Histograma dos valores de ARL somente da parte Shewart do gráfico combinado para o Exemplo 6.1 com $p = 0,09$.

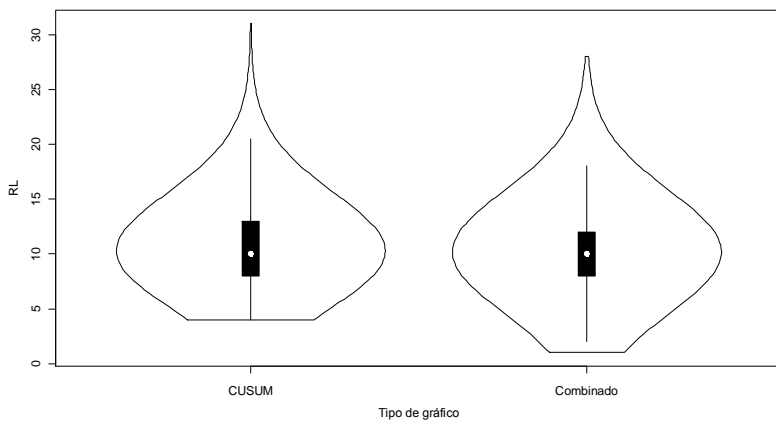


Figura 6. 7 – Gráfico violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e somente da parte CUSUM para o Exemplo 6.1 com $p = 0,09$.

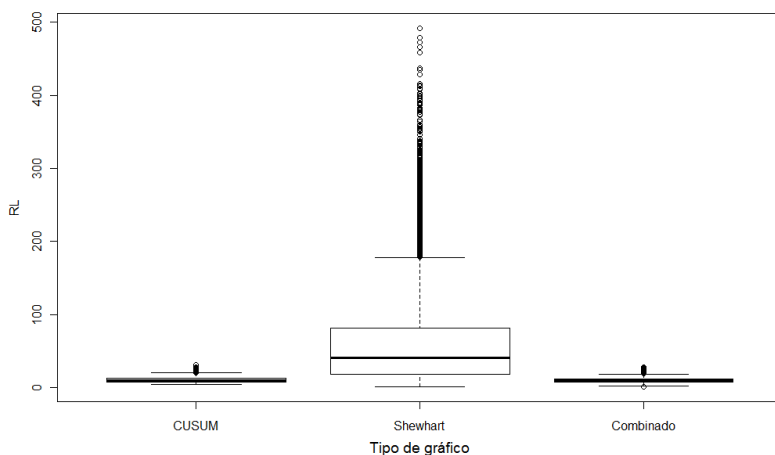


Figura 6. 8 – Gráfico Boxplot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,09$.

Nas Figuras 6.9 a 6.17 esta crescente participação da parte Shewhart na detecção é vista. A distribuição do combinado vai aos poucos “incorporando” a parte Shewhart, sendo esta responsável pelos percentis menores e a parte CUSUM pelos maiores. Não foram plotados histogramas em virtude da amplitude dos valores da parte Shewhart, que tornou complexa a escolha de número de classes que permitisse uma comparação com o combinado e o CUSUM. Todavia, o gráfico violin plot permite uma boa visualização do que acontece.

É interessante verificar o comportamento da distribuição do combinado, quanto à simetria. Para p próximo de p_0 , a distribuição é assimétrica à direita (Figuras 6.1 a 6.4). Na medida em que p aumenta vai se tornando mais simétrica (Figuras 6.5, 6.7, 6.9 e 6.10), e depois, para grandes alterações, volta a ser assimétrica à direita novamente (Figuras 6.13 a 6.17). Os valores dos coeficientes de assimetria (CSRL) estão na Tabela 6.1.

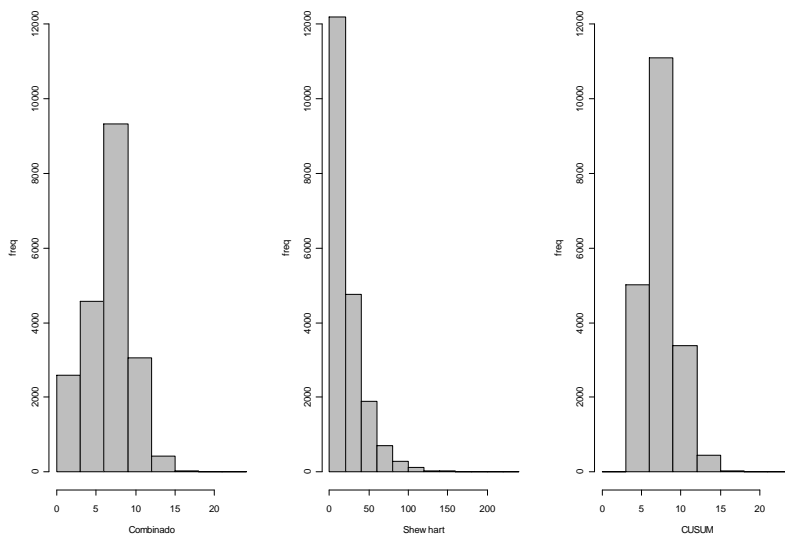


Figura 6. 9 – Histogramas dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,102$.

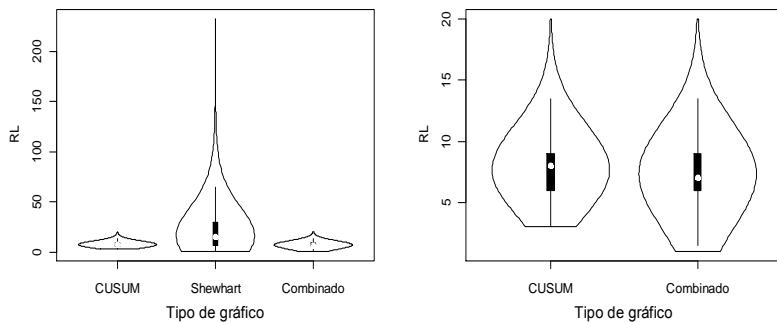


Figura 6. 10 – Gráfico violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais correspondentes para o Exemplo 6.1 com $p = 0,102$.

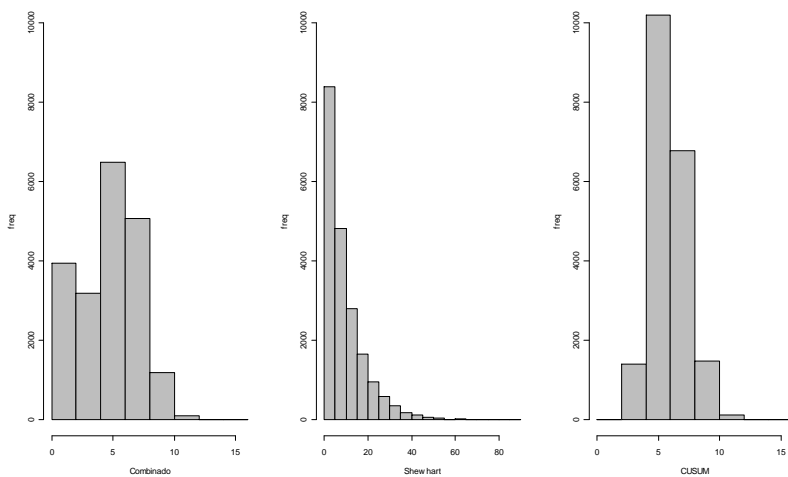


Figura 6. 11– Histogramas dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,114$.

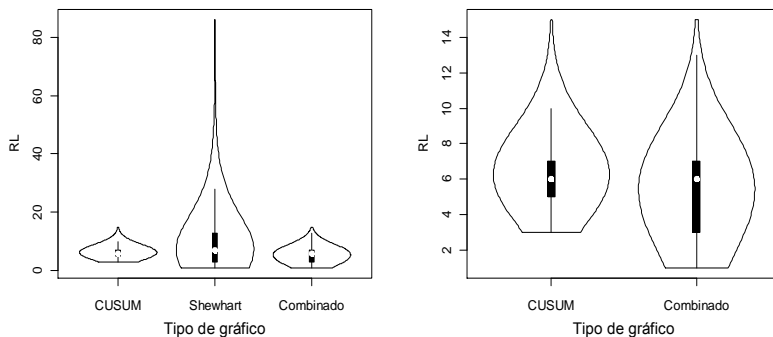


Figura 6. 12 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,114$.

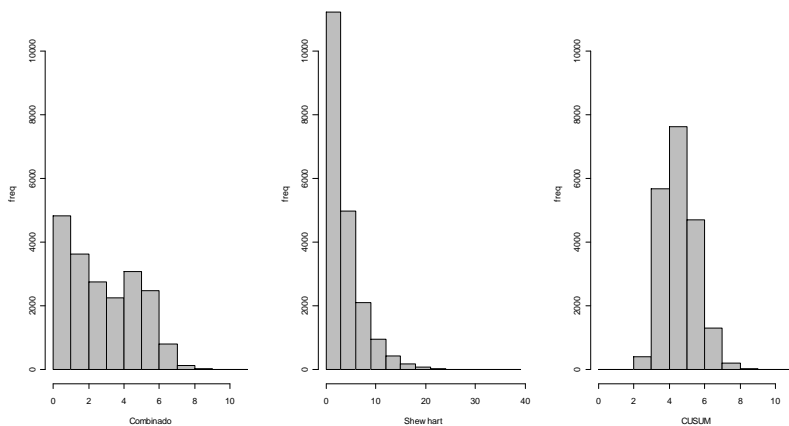


Figura 6. 13 – Histograma dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,120$.

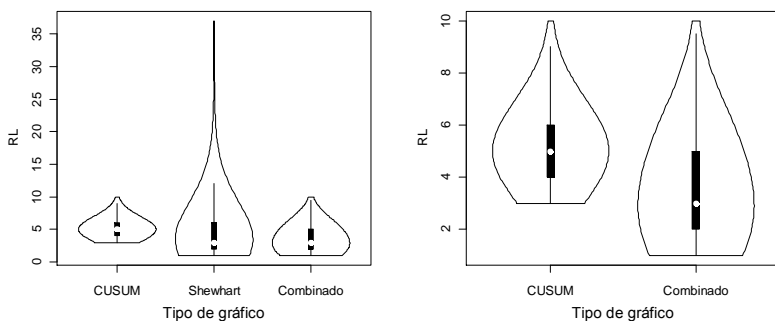


Figura 6. 14 – Gráficos violín plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 comp = 0,120.

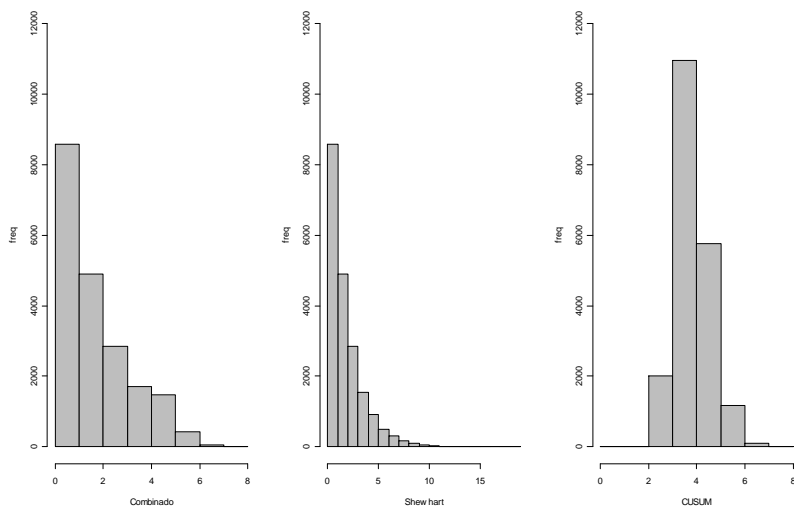


Figura 6. 15– Histogramas dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,150$.

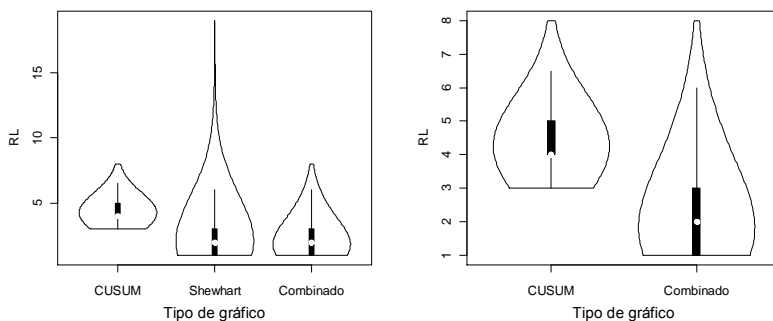


Figura 6. 16 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,150$.

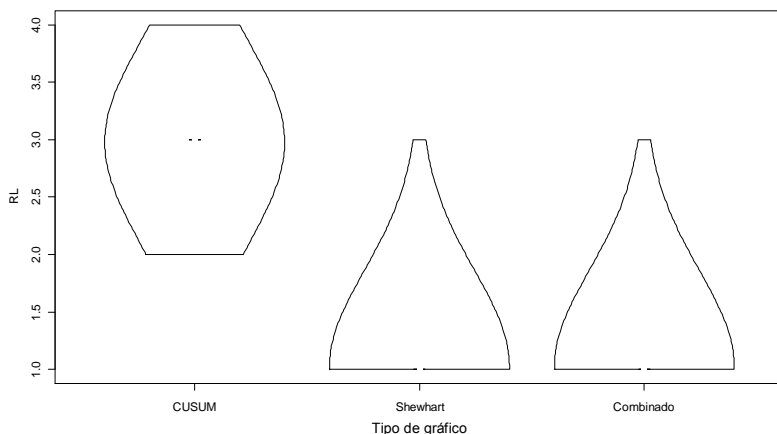


Figura 6. 17 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais (Shewhart e CUSUM) para o Exemplo 6.1 com $p = 0,240$.

6.1.3 RL do gráfico combinado - exemplo 6.2

Outro exemplo é apresentado a seguir. Seja um gráfico combinado para: $p_0 = 0,02$; $p_1 = 1,25p_0 = 0,025$ ($s = 25\%$; $0,25\sigma$); $n = 50$ e $ARL_0 = 200$. Os parâmetros são: parte CUSUM com $ARL_0 = 220$; $h = 8,6$ e $k = 1,12$; parte Shewhart, com $ARL_0 = 2091$, com limite probabilístico (ou exatos) $LSC = 5$. O limite superior de controle (LSC) foi calculado usando a distribuição binomial. Como $ARL_0 = 2091$, a taxa de alarmes falsos $\alpha = 1/2091 = 0,0004782401$, o mesmo procedimento de simulação foi feito para estes dados. Neste exemplo foi possível atingir um valor próximo do ARL_0 de projeto na parte Shewhart do combinado. Os valores de ARL estão na Tabela 6.2 e nas Figuras 6.18 e 6.19 estão os gráficos violin plots para alguns valores de p .

Tabela 6. 2–Valores de ARL, CSRL e CKRL para o gráfico combinado do Exemplo 6.2 e valores de ARL das partes CUSUM e Shewhart.

s	p (%)	Parte		Gráfico		
		CUSUM	Shewhart	COMBINADO	CSRL	CKRL
0	0,020	214,3	2026	204,00	1,92	8,25
10	0,022	98,94	1171	96,14	1,89	8,20
25	0,025	45,79	670,2	44,95	1,88	8,51
50	0,030	22,59	178,1	22,16	1,50	6,97
75	0,035	15,08	117,0	14,69	1,18	5,33
100	0,040	11,05	68,90	11,08	0,87	4,95
125	0,045	9,37	40,92	8,87	0,56	4,21
150	0,050	8,02	26,85	7,46	0,40	3,82
175	0,055	7,05	17,85	6,40	0,23	3,16
200	0,060	6,29	13,00	5,55	0,25	2,97
225	0,065	5,66	9,52	4,79	0,25	2,79
250	0,070	5,25	7,43	4,31	0,24	2,59
300	0,080	4,57	4,81	3,43	0,41	2,55
350	0,090	3,43	4,11	2,79	0,60	2,62
400	0,100	3,77	2,61	2,34	0,80	2,77
650	0,150	2,88	1,28	1,28	2,31	8,64

Na Figura 6.18, valores de p menores, são comparados com os violin plots do combinado e da parte CUSUM. Gráficos da parte Shewhart são omitidos, em virtude da amplitude dos valores, e da preponderância da parte CUSUM na detecção.

Na Figura 6.19, para valores maiores de p , comparam-se as partes CUSUM e Shewhart com o gráfico combinado. Pode-se verificar que o ARL_0 de projeto é alcançado e o comportamento do RL do gráfico combinado é similar ao do Exemplo 6.1.

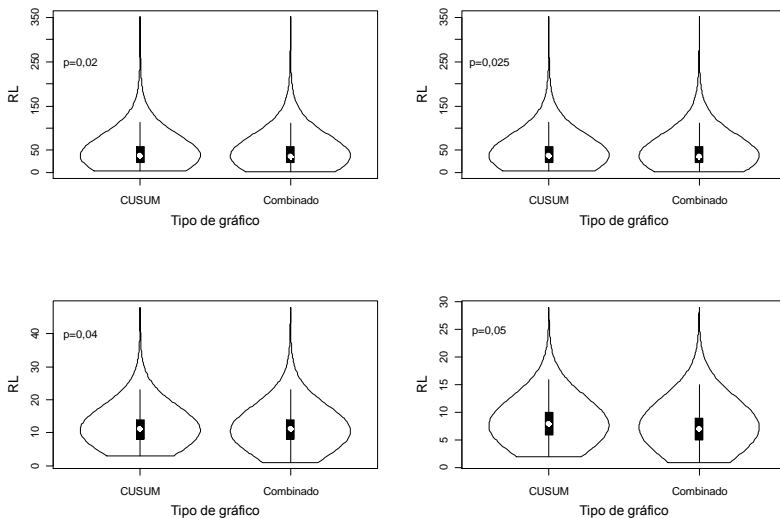


Figura 6. 18 – Gráficos dos valores de ARL do gráfico combinado Shewhart-CUSUM e da parte CUSUM individual para o Exemplo 6.2 com $p = 0,02; 0,025; 0,04$ e $0,05$.

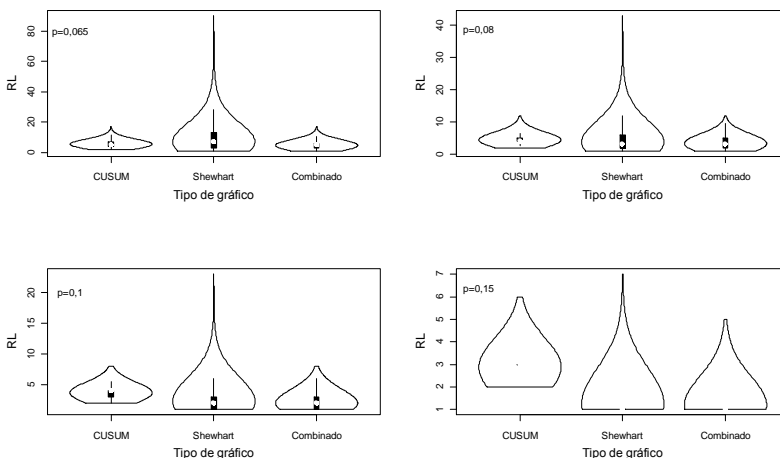


Figura 6. 19 – Gráficos violin plot dos valores de ARL do gráfico combinado Shewhart-CUSUM e das partes individuais Shewhart e CUSUM para o Exemplo 6.2 com $p = 0,065; 0,08; 0,10$ e $0,15$.

6.1.4 Algumas considerações sobre o RL do combinado

O RL de um gráfico combinado para pequenas alterações se comporta como o RL de um gráfico CUSUM. Inicia, quando p igual ou próximo a p_0 , com características de uma distribuição assimétrica e, na medida em que p aumenta, torna-se um pouco mais simétrico. Para valores grandes de p , ele volta a ser assimétrico, com a parte Shewhart preponderante na detecção. A assimetria na distribuição do RL reforça a imperfeição do ARL como medida de desempenho, para o CUSUM, para o Shewhart e também para o combinado. Para o Exemplo 6.2 calculando-se o valor do primeiro quartil (q_2) para a situação sob controle, verifica-se que há uma probabilidade de 0,25 de que um alarme falso seja emitido até a amostra 70 (para ambos), de 0,50 (MRL) até a amostra 148 (para ambos) e de 0,75 (terceiro quartil ou q_2) até as amostras 276 (combinado) e 279 (CUSUM). Os valores de SDRL são altos, (186 e 187), quando o valor da média (ARL_0) é em torno de 200. Na prática, isto quer dizer que o gráfico tem uma probabilidade razoável, que não pode ser desprezada, de sinalizar antes, ou depois, do esperado.

6.2 DESEMPENHO DO GRÁFICO COMBINADO

Nesta seção o desempenho do gráfico combinado será comparado com o de um gráfico CUSUM individual. O objetivo é verificar se a adição das linhas de Shewhart aumenta a sensibilidade do CUSUM. Em alguns exemplos foi possível comparar o desempenho também com o gráfico do tipo Shewhart. Foram aplicados procedimentos de simulação similares ao da seção anterior, inclusive para os gráficos CUSUM e Shewhart. Além dos valores de ARL, foram estimados os valores dos quartis (q_1 , MRL, q_3) e do desvio padrão (SDRL). Duas medidas de comparação de valores de ARL encontradas na literatura, $rARL$ (WU *et al.*, 2008) e redução de ARL (MORAIS e PACHECO, 2006) complementam a análise.

6.2.1 Comparação de desempenho – Exemplo 6.3

Este exemplo trata da comparação do desempenho de um gráfico combinado e de um gráfico CUSUM com os seguintes parâmetros: $p_0 = 0,02, p_1 = 1,25p_0 = 0,025$ ($s = 25\%; 0,25\sigma$), $n = 50$ e $ARL_0 = 200$. O gráfico combinado tem os seguintes parâmetros: parte CUSUM com $ARL_0 = 220, h = 8,6$ e $k = 1,12$; parte Shewhart com $ARL_0 = 2091$ com limite probabilístico (ou exato) $LSC = 5$. O limite superior de controle foi calculado usando a distribuição binomial. Como $ARL_0 = 2091$, a taxa de alarmes falsos $\alpha = 1/2091 = 0,000478$. Assim, calculando-se o valor de LSC , de modo que $P(x > LSC) = \alpha$, com distribuição $x \sim \text{binom}(0,02; 50)$, $LSC = 5$. Com este valor de LSC , é possível alcançar um valor próximo ao ARL_0 desejado, pois $P(x > 5) = 0,000478$ ($\alpha = 2091,1$). Optou-se por não utilizar os limites convencionais de Shewhart, pois $np = 1 < 5$ e $np(1 - p) = 0,98 < 9$. Se fossem utilizados os limites convencionais, $LSC = 3,6, P(x > 3,6) = 0,0177$, então $ARL_0 = 56,31$.

O gráfico CUSUM individual, para comparação, tem parâmetros $k = 1,12$ e $h = 8,3$ ($ARL_0 = 204,36$). Não será feita comparação, para este exemplo, com um gráfico Shewhart, para o mesmo valor de ARL_0 . Não foi possível, usando limites convencionais ou probabilísticos, alcançar o ARL_0 desejado. Para o valor de $LSC = 4$, $ARL_0 = 311,55$ e para $LSC = 3$, $ARL_0 = 56,31$. Na Tabela 6.3 estão os valores do ARL, primeiro quartil (q_1), MRL, terceiro quartil (q_3) e SDRL, para o gráfico combinado e CUSUM. Na Figura 6.20 está um gráfico comparativo dos logaritmos dos valores de ARL (eixos das ordenadas) dos gráficos combinado e CUSUM. A aplicação do logaritmo facilita a visualização. No eixo das abscissas tem-se a proporção p em termos de $d \times p_0$, onde $d = (s/100 + 1)$. Assim, uma alteração de 25% em p_0 é representada como $1,25p_0$ ¹⁵.

Observando-se os valores ARL do gráfico combinado na Tabela 6.3 e o gráfico da Figura 6.20, percebe-se que o gráfico combinado tem um desempenho similar ao CUSUM quando $p = p_1$. O combinado tem desempenho superior ao CUSUM para valores maiores de p . Neste exemplo, o combinado sinaliza antes do CUSUM para valores de $s \geq 75\%$ e mudanças de pelo menos $0,75\sigma$ em relação à média. Para valores de $p < p_1$ o desempenho do CUSUM é um pouco melhor que o do combinado.

¹⁵Esta notação será aplicada nos demais gráficos do Capítulo 6 e nos gráficos do Capítulo 7.

Tabela 6.3 - Valores de ARL, quartis e SDRL para os gráficos combinado (CS) e CUSUM (C) do exemplo 6.3

s	p	q ₁		MRL (q ₂)		q ₃		ARL		SDRL	
		CS	C	CS	C	CS	C	CS	C	CS	C
0	0,020	70	70	148	148	276,2	279	204,00	201,10	187,57	186,82
10	0,022	38	38	72	70	129	126	96,14	94,76	81,74	80,02
25	0,025	22	22	36	36	58	57	44,95	44,52	32,57	31,56
50	0,030	13	13	19	19	28	28	22,16	22,16	12,54	12,26
75	0,035	10	10	13	13	18	18	14,69	14,86	7,23	6,85
100	0,040	8	8	11	11	14	14	11,08	11,34	4,97	4,65
125	0,045	6	7	9	9	11	11	8,87	9,29	3,87	3,50
150	0,050	5	6	7	7	9	10	7,46	7,85	3,27	2,82
175	0,055	5	5	6	6	8	8	6,40	6,91	2,88	2,34
200	0,060	4	5	5	6	7	7	5,55	6,17	2,60	1,99
225	0,065	3	4	5	5	6	6	4,79	5,61	2,36	1,71
250	0,070	2	4	4	5	6	6	4,31	5,18	2,19	1,50
300	0,080	2	4	3	4	5	5	3,43	4,53	1,88	1,18
350	0,090	1	3	2	4	4	5	2,79	4,10	1,61	0,99
400	0,100	1	3	2	4	3	4	2,34	3,76	1,39	0,85
650	0,150	1	3	1	3	1	3	1,28	2,89	0,58	0,55

Analisando o valor da mediana, o desempenho do combinado é similar ao comportamento do ARL (Tabela 6.3 e Figura 6.21). Pode-se perceber que o procedimento combinado tem desempenho similar ao CUSUM individual para alterações menores e, desempenho superior (sinalizando antes) para alterações maiores em p_0 .

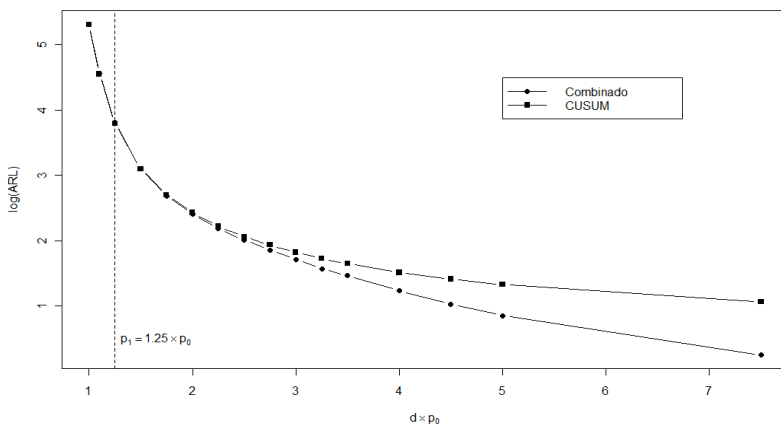


Figura 6. 20– Comparação dos valores de log(ARL) para o exemplo 6.3.

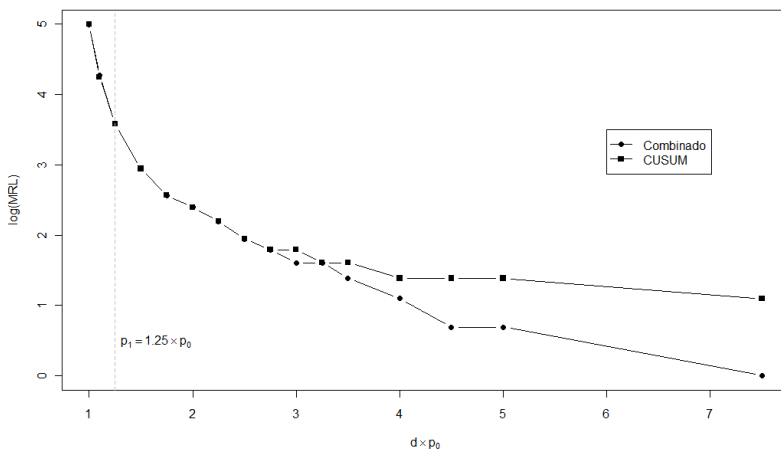


Figura 6. 21 – Comparação dos valores de log (MRL) para o exemplo 6.3.

Em sequência é calculado o $rARL$ (WU; JIAO; LIU, 2008), Equação (6.1), que é uma medida criada para facilitar a comparação dos valores de ARL de diversos gráficos. Com esta medida será comparado o ARL do gráfico combinado em relação ao ARL do CUSUM para um específico valor de p .

$$rARL(p) = \frac{ARL_{comb}(p)}{ARL_{CUSUM}(p)} \quad (6.1)$$

Valores de $rARL$ menores que um (1) implicam que o gráfico combinado é mais eficaz que o CUSUM simples para detectar este particular valor de p . Na Figura 6.22 está um gráfico comparativo com os valores de $rARL$ para o exemplo 6.3. São comparados os valores de ARL do procedimento combinado e de um gráfico CUSUM individual. A linha vertical pontilhada corresponde à proporção $p = p_1$ de projeto. A análise deste gráfico reforça que o desempenho do combinado é superior ao CUSUM, na medida em que a magnitude das alterações é maior.

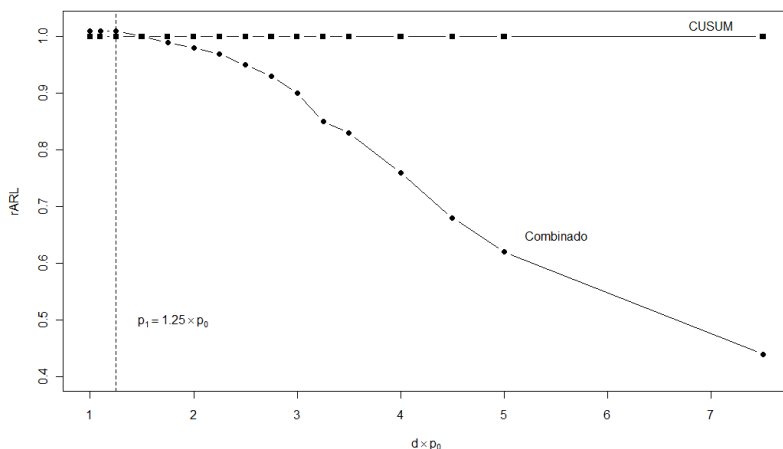


Figura 6. 22 – $rARL$ para o Exemplo 6.3.

Na Figura 6.23 está um gráfico comparativo do $SDRL$ dos gráficos combinados e CUSUM. Verifica-se que com o gráfico combinado a variabilidade, em termos de $SDRL$, de modo geral, tende a ser maior quando comparada a um CUSUM individual. Na prática, isso

significa que o combinado tem probabilidade maior de sinalizar antes, e depois do valor esperado, na medida em que a magnitude de p aumenta. Diminui, embora pouco, para alterações muito grandes, no caso do exemplo, maiores que $5p_0$.

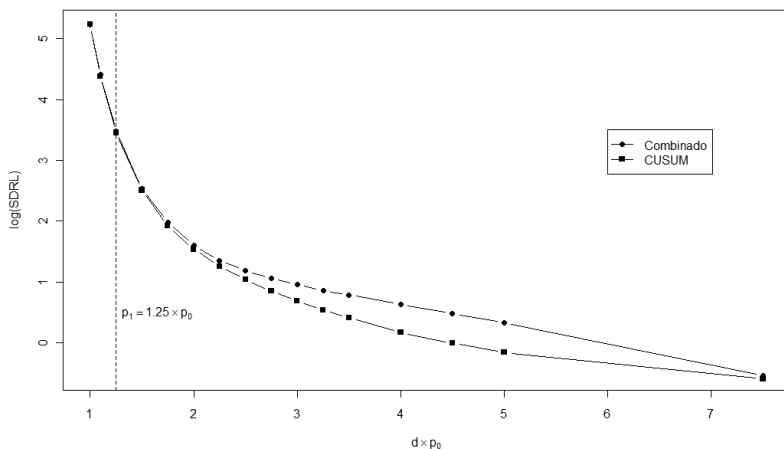


Figura 6. 23 – Comparação de SDRL para o Exemplo 6.3.

6.3 COMPARAÇÃO DE DESEMPENHO – MAIS EXEMPLOS

Os próximos exemplos diferem do anterior em termos de tamanho de amostra, proporção sob controle, s e ARL_0 alvo. O objetivo é verificar se o mesmo comportamento visto no exemplo anterior se repete. São apresentadas tabelas com os valores de ARL , MRL e $SDRL$, gráficos comparativos. As análises são feitas ao final de cada exemplo.

Exemplo 6.4: Seja $p_0 = 0,06$; $n = 100$ e $p_1 = 0,066 = 1,1p_0$ ($s = 10\%$; $0,25\sigma$) e $ARL_0 = 370$. Os parâmetros do gráfico combinado são: $k = 6,3$, $h = 24$ e $LSC = 15$. O ARL_0 da parte é CUSUM é 390. O limite convencional de Shewhart foi utilizado, com $q = 3,8$, $ARL_0(\text{projeto}) = 7196$ aproximadamente. Aqui $np = 6 > 5$, mas $np(1 - p) = 5,65 < 9$. O limite superior do CUSUM individual é 23,5 ($ARL_0 \approx 370$). Neste exemplo será possível comparar o

desempenho com um gráfico Shewhart ($LSC = 13,1$) e um ARL_0 um pouco maior que o de projeto (aproximadamente 383).

A Tabela 6.4 traz os valores de ARL, MRL e SDRL. Na Figura 6.24 está um gráfico comparativo dos valores de ARL.

Em termos de ARL, o combinado tem desempenho superior ao CUSUM para alterações um pouco maiores que as de planejamento, por volta de $s \cong 40\%$. E, em relação ao gráfico do tipo Shewhart para as alterações menores que $s \cong 70\%$ (aproximadamente $1,8\sigma$).

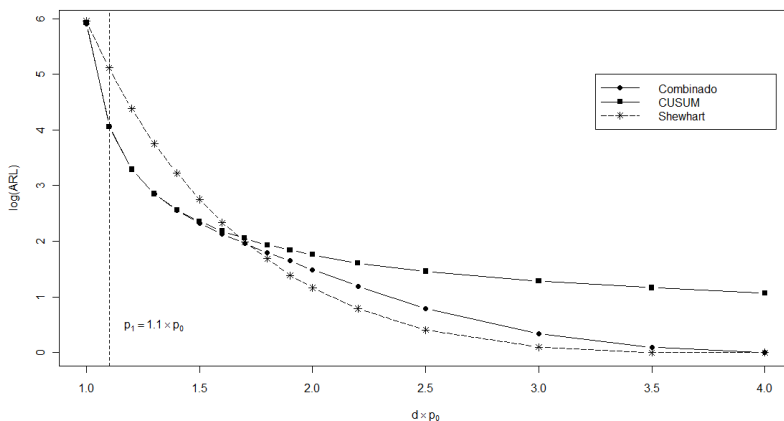


Figura 6. 24– Comparação de $\log(ARL)$, para o Exemplo 6.4.

Percebe-se na Figura 6.24 uma região onde o desempenho do gráfico combinado é superior aos dois gráficos. Esta região pode ser visualizada mais detalhadamente na Figura 6.25. Para alterações entre $s \cong 30\%$ (aproximadamente $0,75\sigma$), até $s \cong 72\%$ (aproximadamente $1,8\sigma$), o gráfico combinado sinaliza antes dos gráficos CUSUM e Shewhart individuais. Esta faixa de valores corresponde a alterações com magnitudes mais moderadas, maiores que as específicas de projeto do CUSUM. O CUSUM individual necessita de mais amostras para a detecção. Com relação ao gráfico Shewhart, a magnitude da mudança ainda não é grande o suficiente para que ele sinalize.

Tabela 6. 4 - ARL, MRL e SDRL, para o exemplo 6.4.

s	p ₀	ARL			MRL			SDRL		
		Comb.	CUSUM	Shewhart	Comb.	CUSUM	Shewhart	Comb.	CUSUM	Shewhart
0	0,060	368,90	374,40	385,80	262,00	265,00	272,00	348,96	353,13	381,60
10	0,066	58,38	57,98	166,16	48,00	48,00	115,00	39,95	38,95	166,16
20	0,072	26,70	26,75	80,03	24,00	24,00	57,00	13,08	12,99	76,63
30	0,078	17,40	17,34	42,58	16,00	16,00	30,00	7,13	6,70	41,74
40	0,084	12,86	12,99	25,08	12,00	12,00	18,00	4,82	4,34	24,23
50	0,090	10,18	10,53	15,66	10,00	10,00	11,00	3,73	3,10	15,24
60	0,096	8,41	8,90	10,27	9,00	9,00	7,00	3,12	2,40	9,76
70	0,102	7,08	7,77	7,26	7,00	8,00	5,00	2,82	1,93	6,68
80	0,108	6,01	6,93	5,35	6,00	7,00	4,00	2,58	1,61	4,77
90	0,114	5,15	6,31	4,02	5,00	6,00	3,00	2,42	1,40	3,47
100	0,120	4,40	5,79	3,19	3,00	6,00	2,00	2,25	1,24	2,63
120	0,132	3,33	5,04	2,23	3,00	5,00	2,00	1,94	0,97	2,23
150	0,150	2,21	4,30	1,53	2,00	4,00	1,00	1,40	0,73	0,91
200	0,180	1,35	3,58	1,13	1,00	4,00	1,00	0,68	0,58	0,39
250	0,210	1,09	3,15	1,03	1,00	3,00	1,00	0,31	0,42	0,17
300	0,240	1,02	2,93	1,01	1,00	3,00	1,00	0,14	0,34	0,07

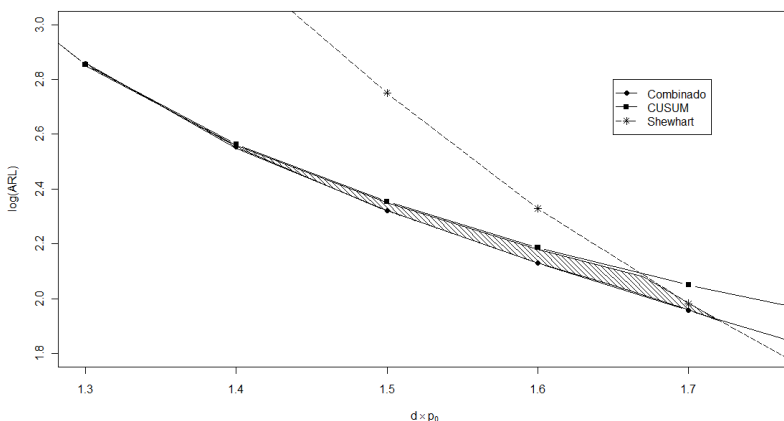


Figura 6. 25 – Região onde o gráfico combinado é mais efetivo que os gráficos CUSUM e tipo Shewhart para o Exemplo 6.4.

Os gráficos comparativos de ARL, MRL, e SDRL (Figuras 6.24, 6.26 e 6.27) do Exemplo 6.4 mostram desempenho similar ao observado no Exemplo 6.3.

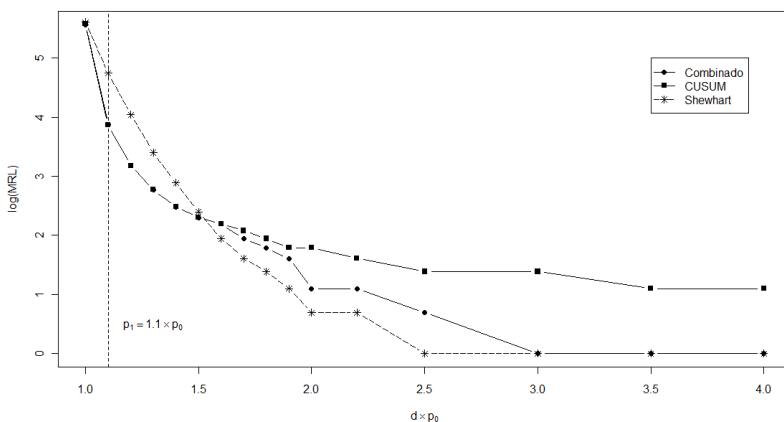


Figura 6. 26 – Comparação de log (MRL), para o Exemplo 6.4.

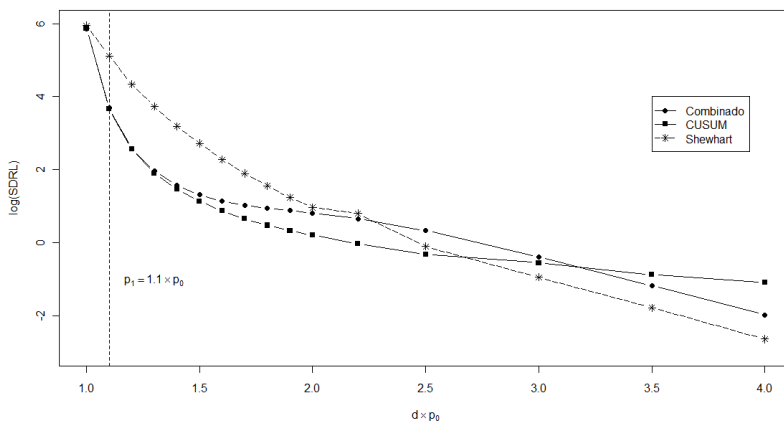


Figura 6. 27 – Comparação de log (SDRL), para o Exemplo 6.4.

Na Figura 6.28 está um gráfico comparativo em termos de rARL entre os gráficos combinado e CUSUM, onde confirma-se que o gráfico combinado tem desempenho superior ao CUSUM para alterações um pouco maiores que as de projeto, neste exemplo para $s > 30\%$. Na Figura 6.29 comparam-se os valores de rARL do gráfico combinado em relação ao gráfico do tipo Shewhart. Pode-se verificar que para alterações de menor magnitude (neste exemplo $s < 75\%$ aproximadamente) o gráfico combinado tem melhor desempenho.

Na Figura 6.30 está a análise em termos de rARL confrontando o gráfico do tipo Shewhart e CUSUM com o combinado, podendo-se visualizar a mesma região onde o combinado tem desempenho superior aos demais (entre $s \cong 30\%$ até $s \cong 72\%$). As análises dos gráficos das Figuras 6.28, 6.29 e 6.30 conduzem aos mesmos resultados das análises das Figuras 6.24 e 6.25.

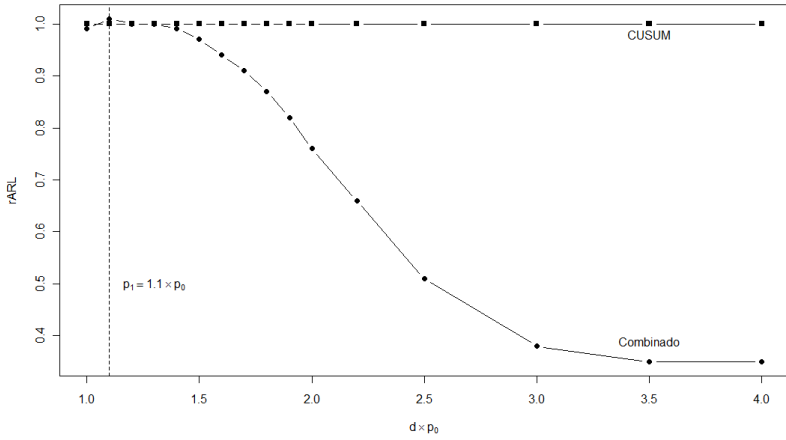


Figura 6. 28 – rARL (CUSUM e combinado), para o Exemplo 6.4.

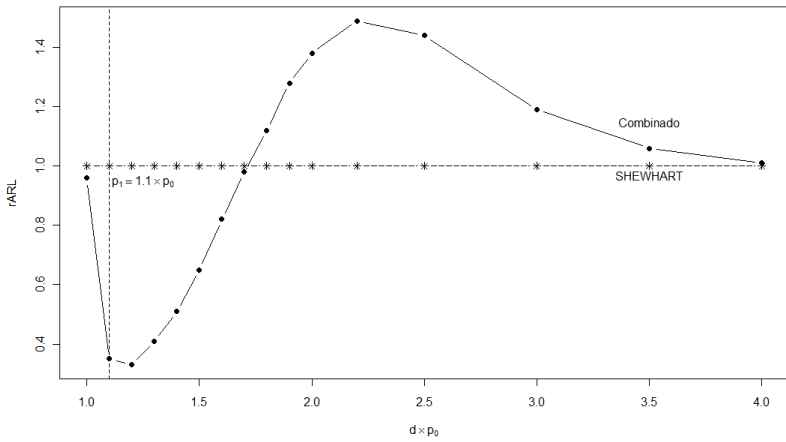


Figura 6. 29 – rARL (Combinado e Shewhart), para o Exemplo 6.4.

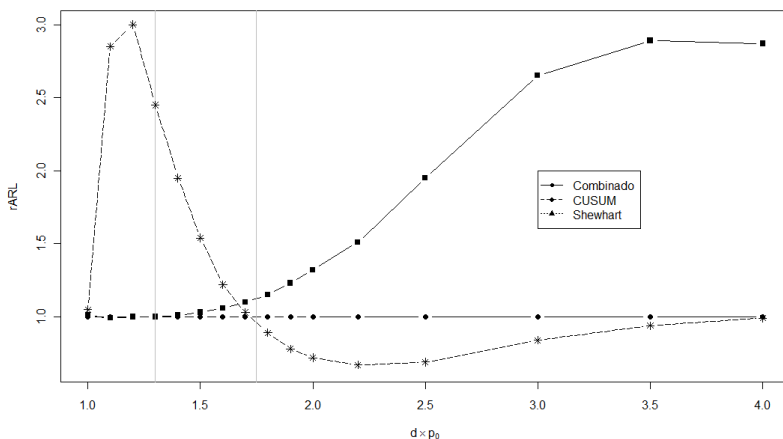


Figura 6. 30 – rARL entre os gráficos CUSUM e tipo Shewhart em relação ao gráfico combinado para o Exemplo 6.4.

Em sequência é feita uma análise, mediante a redução do ARL (*ARL reduction*) (MORAIS e PACHECO, 2006), Equação 6.2, outra medida comparativa de valores de ARL, em que a redução (ou não) destes com o aumento da magnitude da mudança é avaliada.

$$\text{Comb vs. } C = \frac{1 - ARL_{comb}(p)}{ARL_{CUSUM}(p)}. \quad 6.2$$

A Figura 6.31 mostra a redução em termos de ARL do gráfico combinado em relação ao CUSUM e ao gráfico do tipo Shewhart, com formulação análoga a 6.2. Os resultados, em termos de redução de ARL são os mesmos que na análise por rARL (Figuras 6.27 e 6.28).

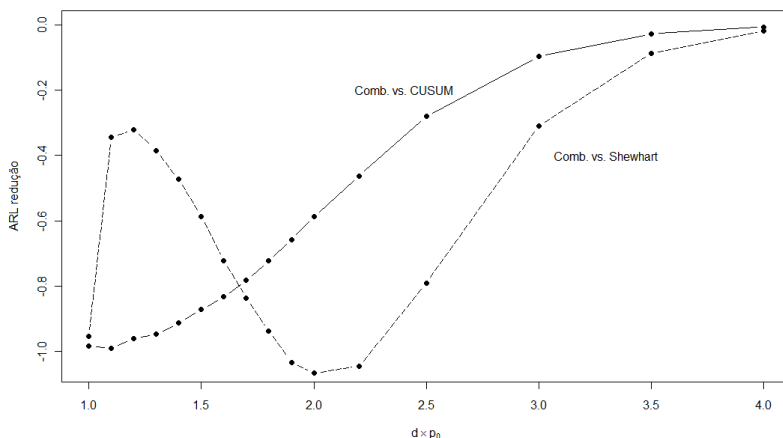


Figura 6. 31– Redução do ARL do gráfico combinado, para o Exemplo 6.4.

Exemplo 6.5: Seja $p_0 = 0,1$, $n = 100$ e $p_1 = 0,11 = 1,1p_0$ ($s = 10\%$; $0,33\sigma$) e $ARL_0 = 50$. Os parâmetros do gráfico combinado são: $k = 10,49$, $h = 12,6$ e $LSC = 19,0$. O ARL_0 da parte de CUSUM é 55,2. Para a parte Shewhart os limites probabilísticos são $LSC = 19$, $ARL_0 = 505$ aproximadamente. Neste exemplo será possível comparar o desempenho com um gráfico Shewhart, com LSC (probabilístico) = 16,0, $ARL_0 \cong 50$. O limite superior do CUSUM individual é $h = 12,5$ ($ARL_0 \cong 50$). A Tabela 6.5 traz os valores de ARL, MRL e SDRL. Na Figura 6.32 está um gráfico comparativo dos valores de ARL. A Figura 6.33 mostra a região onde o gráfico combinado tem desempenho superior aos gráficos individuais. Nas Figuras 6.33 e 6.34 estão as comparações em termos de rARL, e na Figura 6.35 a redução do ARL do combinado em relação aos procedimentos individuais. Os resultados, de modo geral, não diferem dos obtidos nos dois exemplos anteriores, tendo o gráfico combinado desempenho superior para magnitudes de mudança um pouco maiores que as de planejamento do CUSUM, neste exemplo para $s \cong 20\%$. A região onde o gráfico combinado tem desempenho melhor (Figura 6.33) abrange uma faixa entre $s \cong 20\%$ e $s \cong 31\%$ aproximadamente correspondendo a mudanças entre $1,07\sigma$ a $1,5\sigma$ em relação à média (np_0) aproximadamente.

Tabela 6. 5 - ARL, MRL e SDRL, para o Exemplo 6.5.

s	p ₀	ARL			MRL			SDRL		
		Comb	CUSUM	Shewhart	Comb	CUSUM	Shewhart	Comb	CUSUM	Shewhart
0	0,10	54,34	53,04	48,90	39,00	39,00	34,00	48,36	47,33	48,9
10	0,11	17,65	17,14	22,06	14,00	14,00	15,00	12,37	12,03	21,83
20	0,12	9,51	9,55	11,47	8,00	8,00	8,00	5,12	5,09	10,91
30	0,13	6,54	6,66	6,66	6,00	6,00	5,00	3,05	2,84	6,06
40	0,14	5,00	5,30	4,29	5,00	5,00	3,00	2,20	1,89	3,76
50	0,15	4,03	4,48	3,05	4,00	4,00	2,00	1,80	1,37	2,49
60	0,16	3,30	3,91	2,29	3,00	4,00	2,00	1,56	1,07	1,69
70	0,17	2,77	3,55	1,86	3,00	3,00	1,00	1,39	0,89	1,27
80	0,18	2,34	3,27	1,56	2,00	3,00	1,00	1,22	0,76	0,93
90	0,19	1,99	3,05	1,37	2,00	3,00	1,00	1,06	0,68	0,71
100	0,20	1,73	2,87	1,24	1,00	3,00	1,00	0,92	0,62	0,55

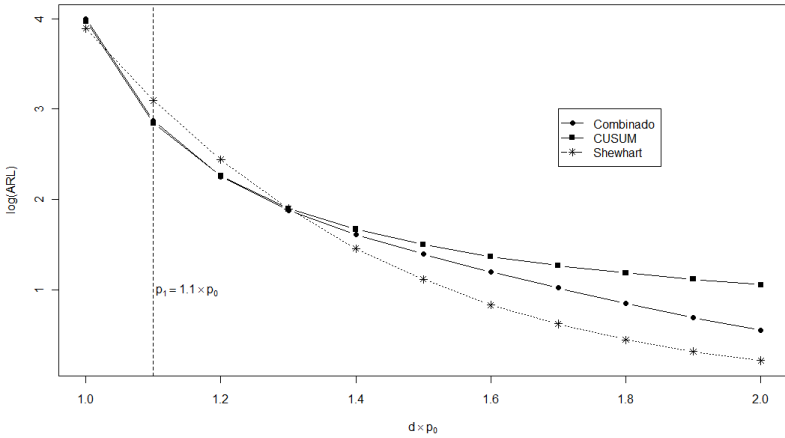


Figura 6. 32– Comparação de valores $\log(\text{ARL})$, para o Exemplo 6.5.

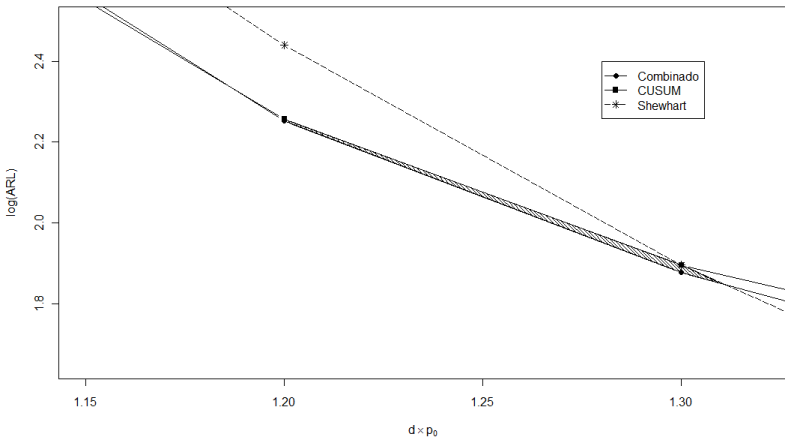


Figura 6. 33– Região onde o gráfico combinado é mais efetivo que os gráficos CUSUM e tipo Shewhart para o Exemplo 6.4.

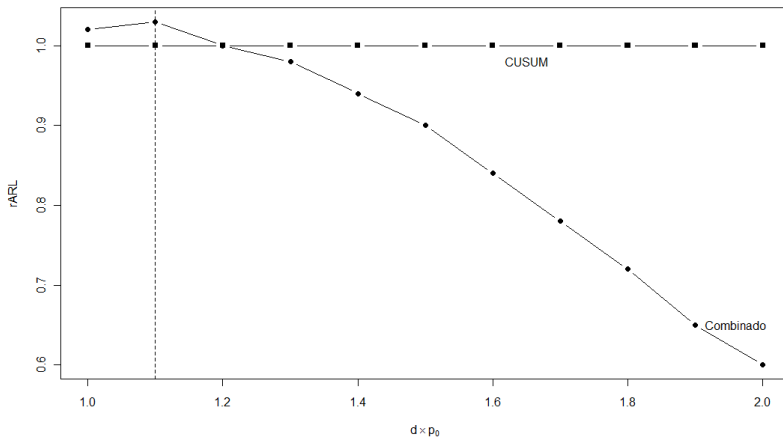


Figura 6. 34– rARL (Combinado e CUSUM) para o Exemplo 6.5.

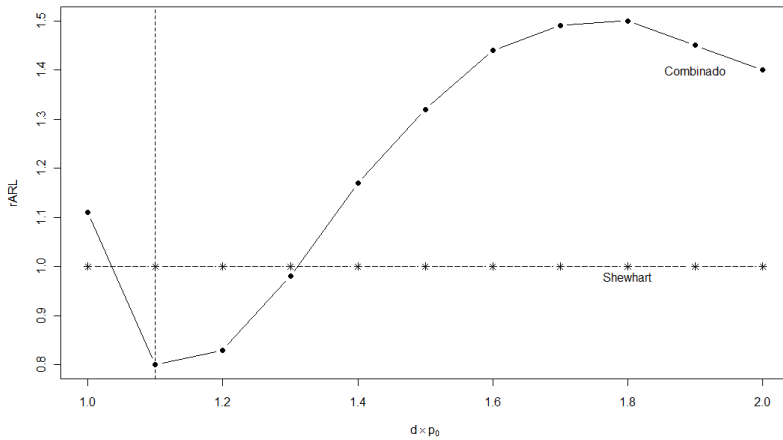


Figura 6. 35 – rARL (Combinado e Shewhart), para o Exemplo 6.5.

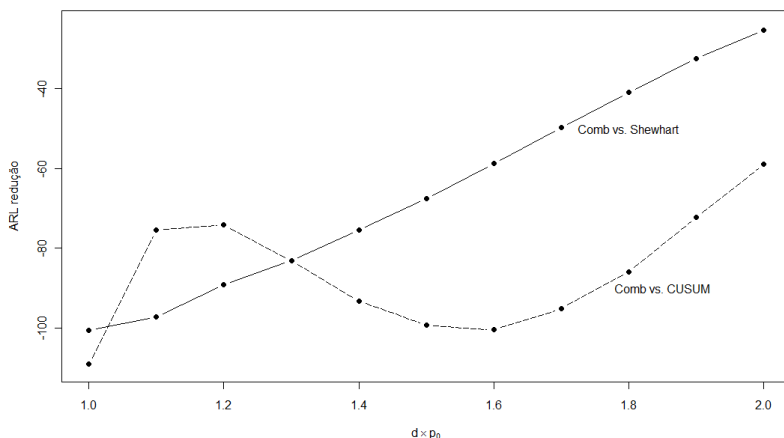


Figura 6. 36 – Redução de ARL, para o Exemplo 6.5.

6.4 SOBRE O DESEMPENHO

Analisando os três exemplos algumas conclusões podem ser realizadas:

a) O gráfico combinado tem desempenho superior ao CUSUM individual para alterações maiores que a proporção p_0 ;

b) Não é simples determinar a partir de que ponto o combinado tem desempenho superior ao CUSUM, mas, de modo geral, para valores de p superiores a p_1 (de planejamento);

c) Na medida em que p se distancia de p_1 , ou seja, as magnitudes de mudança aumentam, o gráfico é mais efetivo que o CUSUM;

d) Para alterações menores, próximas de p_1 (de planejamento), o CUSUM e o combinado podem ter o mesmo desempenho, se o valor do limite superior h de ambos os gráficos forem contíguos;

e) Para alterações iguais ou inferiores a p_1 (de planejamento) o CUSUM o tem desempenho superior ao gráfico combinado;

f) Verifica-se, então, que a adição do limite superior de Shewhart realmente aumenta a sensibilidade do CUSUM binomial para alterações maiores.

g) O gráfico combinado terá desempenho superior ao gráfico tipo Shewhart para pequenas alterações e inferior a este para alterações maiores;

h) Há uma região onde o gráfico combinado tem desempenho melhor que os gráficos CUSUM e tipo Shewhart, sendo que esta região abrange uma faixa de alterações de magnitude mais moderadas;

i) Nos casos analisados esta faixa de valores correspondia a mudanças superiores a p_1 e inferiores a $s = 100\%$;

j) Em termos de números de desvios padrão em relação à média, esta faixa variou de $0,75\sigma$ a $1,8\sigma$, conforme o caso.

Os resultados estão de acordo com o encontrado por Lucas (1982), para variáveis, com um pequeno aumento no valor de h , o gráfico combinado tem performance similar ao CUSUM individual e sinaliza grandes alterações antes deste. Tais resultados estão parcialmente de com os resultados de Morais e Pacheco (2006). Um gráfico combinado sinaliza antes do CUSUM alterações maiores e antes do Shewhart, pequenas a moderadas. A diferença entre este estudo e o de Morais e Pacheco (2006) reside em que o ARL_0 de projeto é o mesmo para os esquemas combinados, CUSUM e Shewhart, sendo os parâmetros h e LSC ajustados para alcançar estes valores. Assim, as conclusões são em termos do mesmo ARL_0 . Morais e Pacheco calcularam o valor do ARL para parâmetros (h e LSC) pré-fixados, chegando à conclusão de que o combinado sinaliza antes, mas à custa de mais alarmes falsos, na parte Shewhart, o que é verdade.

Ressalta-se que o desempenho do combinado pode eventualmente ser superior ao do Shewhart (o que aconteceu), mas não é este o foco principal da análise realizada. O que se deseja é um gráfico CUSUM mais sensível a uma faixa maior de alterações, projetado para detectar pequenas alterações e com um bom desempenho para alterações maiores. Embora o gráfico Shewhart tenha desempenho superior para grandes alterações, existem muitas limitações, abordadas neste trabalho. O balanceamento dos valores de ARL_0 (e parâmetros) de projeto pode auxiliar a contornar o problema de não atingir o valor de ARL_0 , até como uma opção ao gráfico do tipo Shewhart individual.

Lembra-se que nos gráficos comparados neste capítulo a parte CUSUM do combinado foi planejada para detectar pequenas alterações, em que o valor de h espera-se que não seja distante do limite superior h de um CUSUM individual. O limite superior da parte Shewhart é definido para captar as alterações maiores. O balanceamento do ARL_0 é

feito para que o combinado seja o mais próximo possível do CUSUM individual e, assim, verificar se houve aumento da sensibilidade. Isto, é claro, dentro do que a parte Shewhart permite, pois os limites utilizados, convencionais ou probabilísticos, não alcançam sempre os valores de ARL_0 de projeto.

6.5 UMA TABELA COM PARÂMETROS PARA O GRÁFICO COMBINADO SHEWHART-CUSUM BINOMIAL

Para finalizar este capítulo apresentam-se as Tabelas 6.6 e 6.7 com valores dos parâmetros k , h e LSC, para gráficos combinados unilaterais superiores com $s = 20\%$; $ARL_0 = 370$; $n = 50$ (Tabela 6.6) e $n = 100$ (Tabela 6.7). Os limites usados foram os probabilísticos. Procurou-se manter igual ou um pouco superior o valor de h da parte CUSUM do combinado (diferença de um décimo – 0,1) em relação a um CUSUM individual. Estas tabelas são apenas exemplos do que pode ser construído com a metodologia e rotinas desenvolvidas. Mas são muitas as situações e especificidades, como o tamanho da amostra n , os valores de ARL_0 , as proporções p_0 sob controle e valores de s (para definir p_1).

Tabela 6. 6 - Parâmetros de um gráfico combinado com $n = 50$; $s = 20\%$ e $ARL_0 = 370$.

p_0	k	h	LSC
0,01	0,55	9,0	5,0
0,02	1,10	11,1	7,0
0,03	1,65	12,4	8,0
0,04	2,19	13,5	10,0
0,05	2,74	14,1	11,0
0,06	3,29	14,6	12,0
0,07	3,84	15,0	13,0
0,08	4,39	15,2	14,0
0,09	4,94	15,5	15,0
0,10	5,49	15,7	16,0

Tabela 6. 7 - Parâmetros de um gráfico combinado com $n = 50$; $s = 20\%$ e $ARL_0 = 370$.

p_0	k	h	LSC
0,01	1,10	11,2	7
0,02	2,19	13,8	9
0,03	3,29	15,0	12
0,04	4,39	15,8	14
0,05	5,49	16,3	16
0,06	6,58	16,9	17
0,07	7,68	17,2	19
0,08	8,78	17,5	21
0,09	9,88	17,6	22
0,10	10,97	17,7	24

6.6 SÍNTESE DO CAPÍTULO

Neste capítulo analisou-se o comportamento do RL de um gráfico combinado a partir de simulações. A análise de resultados medidas descritivas e gráficas foram aplicadas. Concluiu-se que o RL do combinado é similar ao de um CUSUM para pequenas alterações e ao de um de gráfico Shewhart para grandes alterações. A distribuição é fortemente assimétrica e leptocúrtica para pequenas e grandes alterações, com valores de SDRL altos. Na segunda parte do capítulo o desempenho do gráfico combinado foi comparado com o gráfico CUSUM individual e, a exemplo do Shewhart individual. Concluiu-se que com a adição de linhas de Shewhart é possível aumentar a sensibilidade de um CUSUM para detecção de mudanças maiores. Identificou-se também uma região onde o gráfico combinado tem desempenho superior ao gráfico CUSUM e ao gráfico do tipo Shewhart.

Para complementar, o ARL_0 de projeto do gráfico combinado foi alcançado nas simulações. Não há na literatura parâmetros de ARL_1 para comparação. Os resultados alcançados, para $p \neq p_0$ estão de acordo com um único estudo de caso (MORAIS, 2002; MORAIS e PACHECO, 2006), levando em conta as diferenças entre os estudos.

7. COMPARAÇÃO DO GRÁFICO COMBINADO COM PROCEDIMENTOS CUSUM E SHEWHART PARA SINALIZAR GRANDES ALTERAÇÕES

No Capítulo 6 verificou-se que a adição de linhas de Shewhart aumenta a sensibilidade de um gráfico CUSUM binomial para alterações maiores. Agora o gráfico combinado será comparado com gráficos CUSUM planejados para sinalizar mudanças de amplitudes mais abrangentes, de moderadas, $s = 50\%$, até maiores, como $s = 300\%$. Para complementar, quando possível, estes gráficos serão também comparados com o gráfico Shewhart.

7.1 PROCEDIMENTOS METODOLÓGICOS

Novamente para o procedimento do gráfico combinado são feitas simulações. Já para os gráficos CUSUM os valores de ARL são calculados por Cadeias de Markov¹⁶. A análise é feita apenas em termos de ARL, estes que são comparados graficamente e, também a partir dos valores de $rARL$, visto no Capítulo 6. Os valores de ARL do gráfico do tipo Shewhart foram obtidos também por simulação. São dois os exemplos apresentados neste Capítulo.

7.2 EXEMPLO 7.1

Neste exemplo será comparado o desempenho do gráfico combinado Shewhart-CUSUM com gráficos CUSUM projetados para detecção de valores de p_1 maiores. O gráfico combinado foi desenhado de modo a detectar rapidamente um valor de $p_1 = 1,25p_0$, ou seja, um acréscimo de $s = 25\%$ no valor da proporção sob controle, com $k = 6,3$, $h = 24$ e $LSC = 15$. Os gráficos CUSUM individuais têm valor de referência k calculado para detectar proporções p_1 iguais a $1,25p_0$; $1,5p_0$; $2p_0$ e $3p_0$. Considere-se $p_0 = 0,02$; $n = 50$ e $ARL_0 = 200$. Os parâmetros dos gráficos analisados são:

¹⁶Funções $arICUSUM$, $findH$ e $findK$ do pacote *surveillance* (HOHLE, 2007) no R.

- a) Combinado: $s = 25\%$; $k = 1,12$; $h = 8,6$ ($ARL_0 = 220$) e $LSC = 5$ ($ARL_0 = 2091$);
 b) CUSUM: $s = 25\%$; $k = 1,12$ e $h = 8,3$;
 c) CUSUM 1.5: $s = 50\%$; $k = 1,23$ e $h = 6,4$;
 d) CUSUM 2.0: $s = 100\%$; $k = 1,40$ e $h = 5,0$;
 e) CUSUM 3.0: $s = 200\%$; $k = 1,83$ e $h = 3,2$.

Na Tabela 7.1 estão os valores de ARL relativos aos gráficos combinado e CUSUM acima especificados.

Tabela 7.1 - Valores de ARL, para o Exemplo 7.1.

d	Comb.	CUSUM	CUSUM1.5	CUSUM2.0	CUSUM3.0
1,00	204,00	204,37	203,77	208,4	211,70
1,10	97,16	94,44	99,19	109,33	126,70
1,25	44,16	43,86	45,27	50,98	65,75
1,50	21,11	21,27	20,36	21,51	28,11
1,75	13,51	13,90	12,67	12,48	15,14
2,00	9,96	10,36	9,18	8,62	9,60
2,25	7,85	8,27	7,23	6,58	6,82
2,50	6,41	6,89	5,98	5,34	5,23
2,75	5,45	5,90	5,11	4,51	4,24
3,00	4,71	5,17	4,47	3,92	3,57
3,25	4,17	4,61	3,98	3,48	3,10
3,50	3,72	4,18	3,59	3,13	2,75
4,00	3,01	3,54	3,01	2,63	2,28
4,5	2,54	3,09	2,61	2,28	1,97
5,0	2,17	2,76	2,31	2,03	1,76
7,5	1,26	1,89	1,55	1,38	1,22

Na Figura 7.1 estão os gráficos comparativos de $\log(ARL)$ ¹⁷ do gráfico combinado e CUSUM individual para $p_1 = 1,25p_0$. Nela observa-se que o gráfico CUSUM e o combinado têm o mesmo desempenho para alterações menores e, na medida em que estas aumentam o combinado tem desempenho superior ao CUSUM, como nas análises realizadas no Capítulo 6.

¹⁷A aplicação do logaritmo melhora a visualização do gráfico, conforme Capítulo 6.

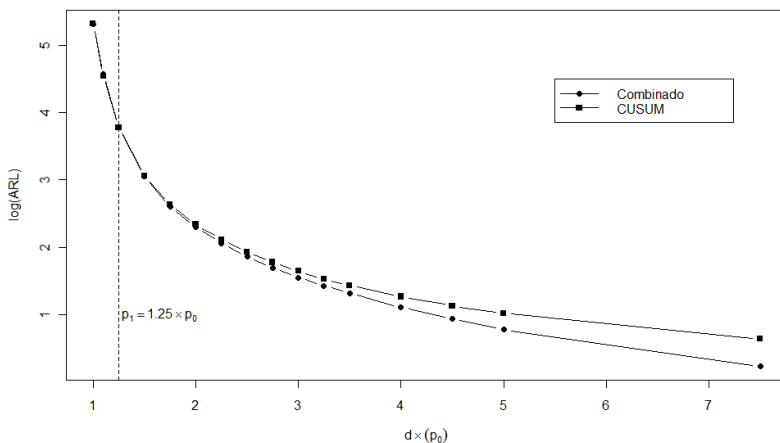


Figura 7. 1 – Comparação de log (ARL) dos gráficos Combinado e CUSUM, para o Exemplo 7.1.

A Figura 7.2 traz os gráficos combinados e CUSUM 1.5 para $p_1 = 1,5p_0$ ($s = 50\%$). Para alterações moderadas, nas proximidades de p_1 do CUSUM individual, este tem desempenho superior. O combinado só tem desempenho superior para alterações superiores a $p_1 = 4p_0$ e nas adjacências de $p_1 = 1,25p_0$.

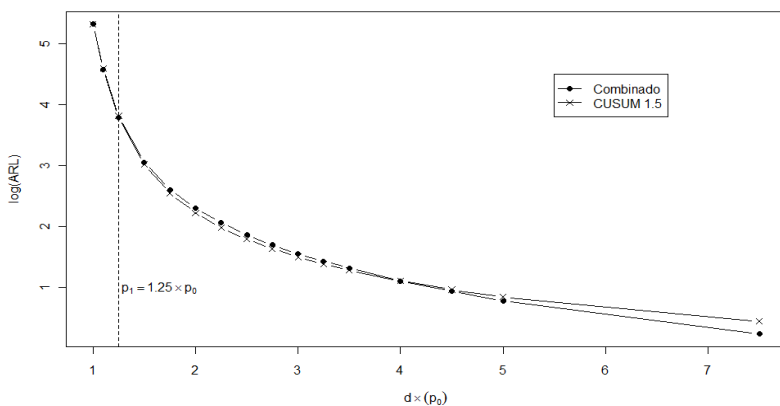


Figura 7. 2 – Comparação de log (ARL) dos gráficos Combinado e CUSUM 1.5, para o Exemplo 7.1.

Na Figura 7.3 estão os gráficos combinados e CUSUM para $p_1 = 2p_0$ ($s = 100\%$). O gráfico combinado é mais efetivo somente para alterações pequenas, $p_1 \leq 1,5p_0$ e superiores e muito grandes, $> 6p_0$. A Figura 7.4 traz os gráficos combinados e CUSUM para $p_1 = 3p_0$ ($s = 200\%$). O CUSUM individual tem desempenho superior para alterações superiores a $p_1 = 2p_0$, ou seja, alterações que podem ser consideradas grandes. O combinado só é mais sensível para as alterações menores, $p_1 < 2p_0$, neste caso.

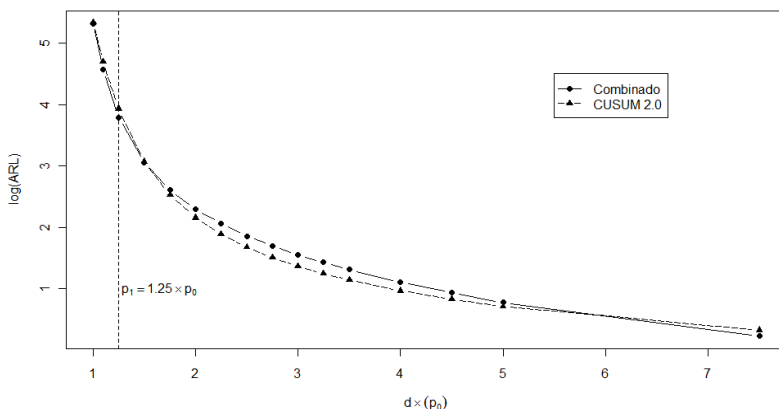


Figura 7.3 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 2.0, para o Exemplo 7.1.

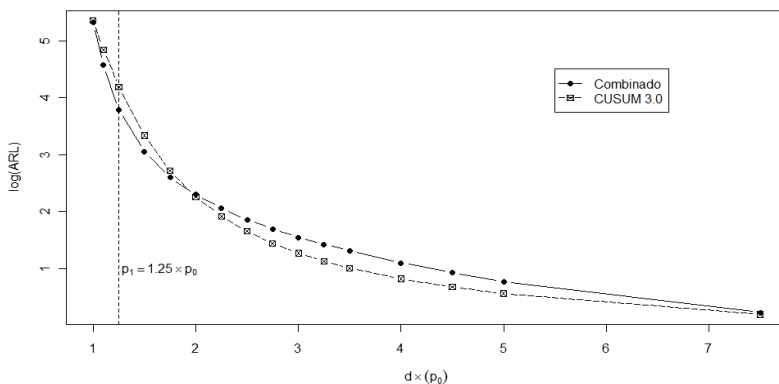


Figura 7.4 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM 3.0, para o Exemplo 7.1.

Complementando, todos os gráficos em conjunto estão na Figura 7.5.

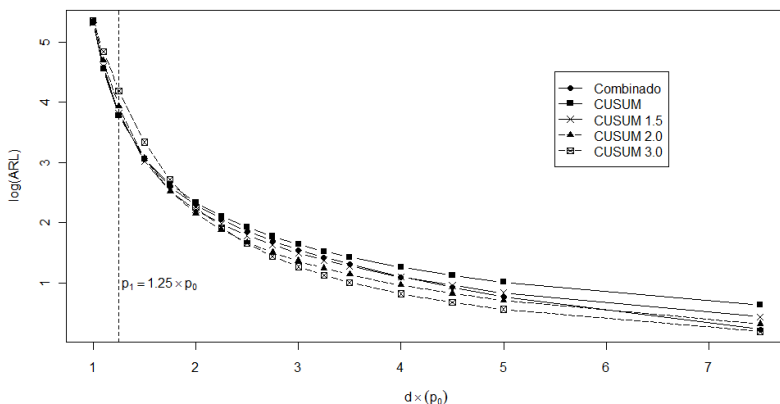


Figura 7.5 – Comparação de $\log(\text{ARL})$ dos Gráficos Combinado e procedimentos CUSUM, para o Exemplo 7.1.

As Figuras 7.6 e 7.7 trazem a análise de $r\text{ARL}$. O ARL do combinado é comparado com o do procedimento CUSUM individual. Por meio destes gráficos comparativos reforçam-se as conclusões anteriores. O desempenho do combinado na detecção de alterações maiores é inferior quando confrontado a procedimentos CUSUM planejados para detecção destas mudanças (Figura 7.6 e 7.7).

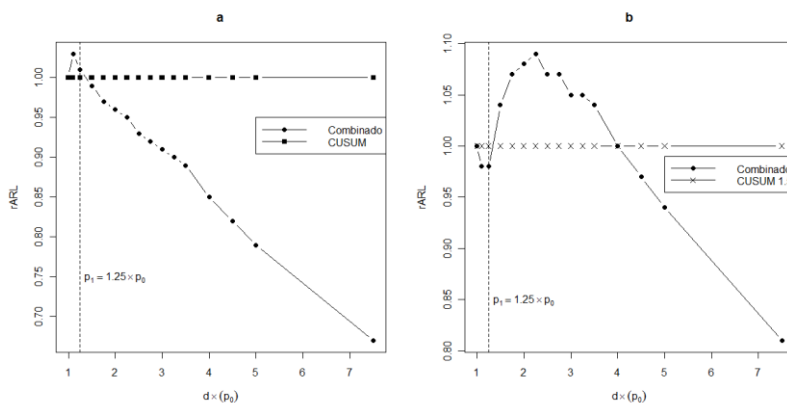


Figura 7.6 – Comparação de $r\text{ARL}$ para os gráficos combinado, CUSUM e CUSUM 1.5, do Exemplo 7.1.

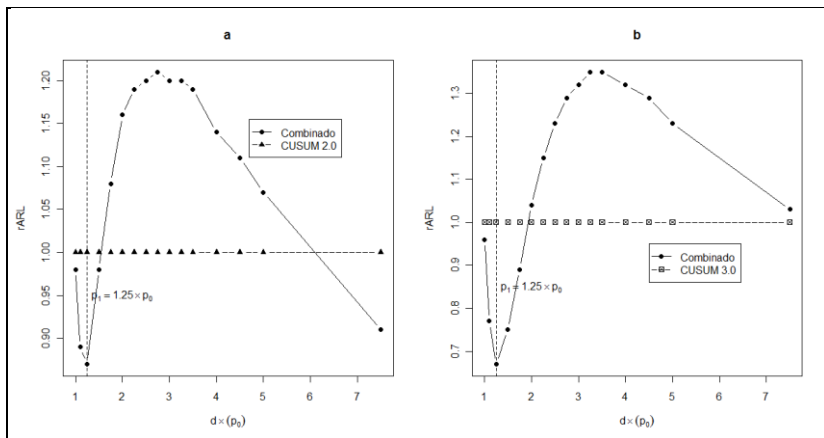


Figura 7.7– Comparação de rARL para os gráficos combinado e CUSUM 2.0 e 3.0, do Exemplo 7.1.

7.3 EXEMPLO 7.2

No próximo exemplo o gráfico combinado foi desenhado de modo a detectar rapidamente um valor de $p_1 = 1,1p_0$, ou seja, um acréscimo de $s = 10\%$ no valor da proporção sob controle. Os gráficos CUSUM individuais têm valor de referência k calculado para detectar proporções p_1 iguais a $1,1p_0$; $1,5p_0$ e $2p_0$. Considere-se $p_0 = 0,1$, $p_1 = 0,11$; $n = 100$ e $ARL_0 = 50$. Os parâmetros dos gráficos são:

- Combinado: $s = 10\%$; $k = 10,49$; $h = 12,6$ ($ARL_0 \cong 55$) e $LSC = 19$ ($ARL_0 = 505$);
- CUSUM: $s = 10\%$; $k = 10,49$ e $h = 12,5$;
- CUSUM 1.5: $s = 50\%$; $k = 12,35$ e $h = 5,0$;
- CUSUM 2.0: $s = 100\%$; $k = 14,52$ e $h = 2,4$ e
- SHEWHART: $LSC = 16,0$ (probabilístico).

Na Tabela 7.2 estão os valores de ARL obtidos e, na Figura 7.8 estão gráficos comparativos do log (ARL) para os procedimentos combinado e apenas CUSUM (CUSUM 1.1) com especificações acima. Neste caso, o valor do LSC do combinado é 19 e o valor do h é 12,6. Aqui $LSC > h$, mas inferior a $k + h$, situação ainda não abordada nas análises anteriores.

Tabela 7. 2 - Valores de ARL para o Exemplo 7.2.

d	comb	CUSUM1.1	CUSUM1.5	CUSUM2.0	Shewhart
1,0	53,29	52,08	50,27	47,50	48,9
1,1	16,58	16,31	19,01	21,21	22,06
1,2	8,69	8,52	9,12	10,93	11,47
1,3	5,77	5,68	5,33	6,37	6,66
1,4	4,26	4,28	3,61	4,12	4,29
1,5	3,39	3,46	2,70	2,92	3,05
1,6	2,78	2,92	2,16	2,22	2,29
1,7	2,34	2,54	1,82	1,80	1,86
1,8	2,02	2,26	1,59	1,53	1,56
1,9	1,77	2,05	1,42	1,35	1,37
2,0	1,56	1,87	1,3	1,23	1,24

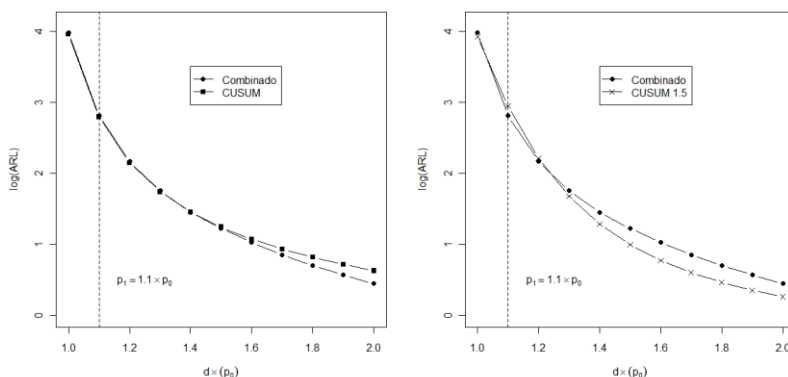


Figura 7. 8 – Comparação de log (ARL) dos gráficos Combinado e CUSUM 1.1e 1.5, para o Exemplo 7.2.

Percebe-se que o gráfico combinado tem desempenho similar ao CUSUM (10%) para pequenas e moderadas alterações. O gráfico combinado é mais efetivo para alterações maiores, para valores de s superiores a 50%, aproximadamente (Tabela 7.2 e Figura 7.8a). Os procedimentos CUSUM planejados para detecção de mudanças maiores ($s = 50$ e 100%) são mais efetivos que o gráfico combinado, especificamente para estas mudanças (Tabela 7.2 e Figuras 7.8 e 7.9).

Na Figura 7.9, além de comparar o gráfico combinado com o CUSUM 2.0 ($s = 100\%$) estes são comparados com um procedimento Shewhart, em termos de ARL. Verifica-se que o gráfico Shewhart tem desempenho similar ao do CUSUM planejado para detectar um aumento de 100% na média. Para complementar, na Figura 7.10 estão os gráficos comparativos entre o gráfico combinado e os procedimentos CUSUM (1.1, 1.5 e 2.0) em conjunto. No gráfico (Figura 7.10), verifica-se que, para este exemplo, o gráfico combinado tem desempenho melhor que um CUSUM individual para alterações de maior magnitude, quando confrontado com o gráfico CUSUM com a mesma proporção p_1 de planejamento. Em relação aos outros gráficos CUSUM, planejados para detectar alterações maiores, seu desempenho é inferior, na faixa de valores analisada.

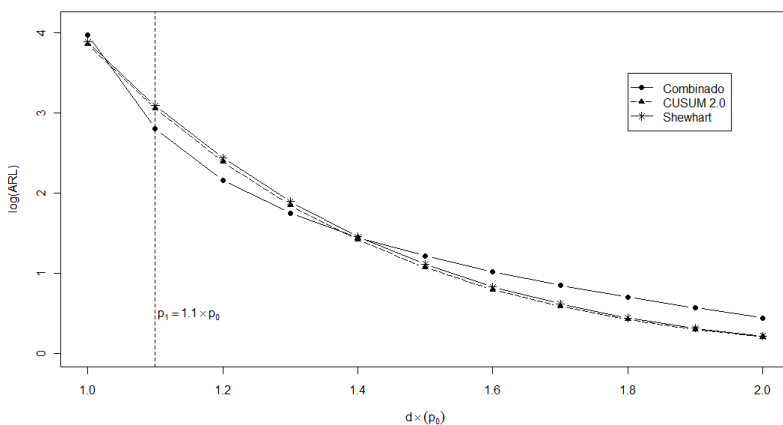


Figura 7.9 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado, CUSUM 2.0, e Shewhart para o Exemplo 7.2.

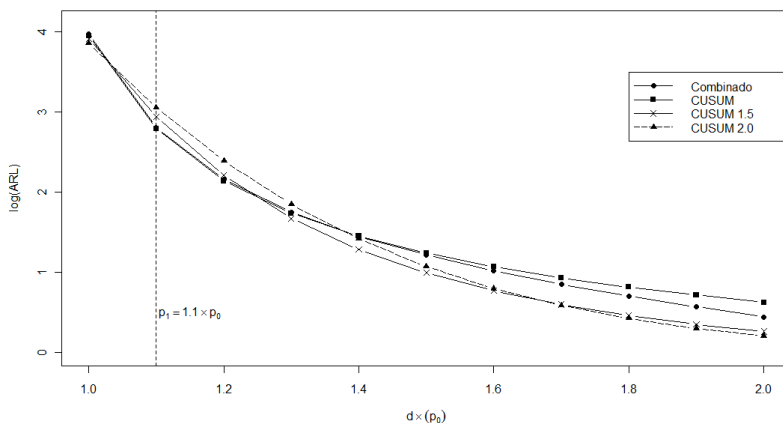


Figura 7. 10 – Comparação de $\log(\text{ARL})$ dos gráficos Combinado e CUSUM, CUSUM 1.5 e CUSUM 2.0 para o Exemplo 7.2.

7.4 COMPARAÇÃO COM O USO DE RIR (RESPOSTA INICIAL RÁPIDA)

Pensou-se em complementar as comparações realizadas nas seções anteriores com o uso do procedimento de Resposta Inicial Rápida (RIR). Para melhorar a sensibilidade do gráfico CUSUM no início, Lucas e Crosier (1982) propõem a implementação deste procedimento cuja utilização permite a redução do valor de ARL_0 e uma redução significativa do ARL_1 (ALVES, 2009). Quando um processo começa fora de controle o uso do RIR permite ao CUSUM detectar isso mais rapidamente, resultando em valores de ARL_1 menores (MONTGOMERY, 2004). Mas, para este estudo, em que se fixou o valor de ARL_0 , para um CUSUM com RIR, o valor de h teria de ser planejado para tal. Apenas considerar um valor diferente de zero ($h/2$, por exemplo) para C_0 , levaria a emissão de mais alarmes falsos. Para esclarecer podem ser verificadas as Figuras 7.11 e 7.12, em que os parâmetros k e h do CUSUM são calculados com o programa *anygeth* (HAWKINS, 1998). Pode-se perceber que o valor de h é menor para um CUSUM sem RIR ($h = 18,2$) do que com RIR ($h = 19,0$). O mesmo

procedimento é feito no R, Figura 7.13, retornando aos mesmos resultados.

```

Enter the sample size and in-control probability : 100 0.05
and the out-of-control probability : 0.056
The exact theoretical reference value is      5.295
Enter the reference value you want to use : 5.3
What are the Winsorizing constants? <say -999 999 if you don't want
to winsorize or don't understand question>: -999 999
Want zero-start (say Z), steady state (say S) or FIR (say F)? : z
Enter ARL : 240
h  18.5000 arls  257.7   223.2   243.3
h  18.3000 arls  249.7   216.0   235.7
h  18.2000 arls  245.9   212.2   232.0
h  18.1000 arls  242.2   209.3   228.4
k  5.3000 h 18.0000 ARL  238.29
      h 18.1000 ARL  242.16
DI 18.100. IC ARL  242.2. OOC ARL  43.1. FIR ARL  29.2. SS ARL  31.7

```

Figura 7. 11 – Saída do programa anygeth para ARL sem FIR.
 FONTE: Programa anygeth¹⁸ (HAWKINS e OLWELL, 1998).

```

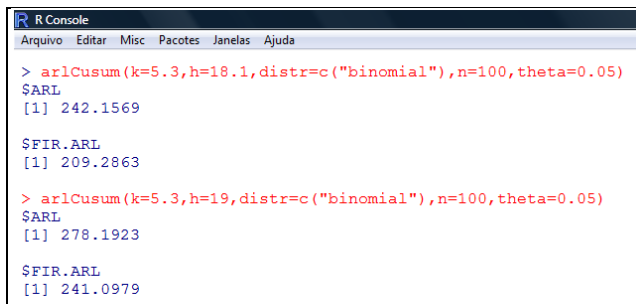
Enter the sample size and in-control probability : 100 0.05
and the out-of-control probability : 0.056
The exact theoretical reference value is      5.295
Enter the reference value you want to use : 5.3
What are the Winsorizing constants? <say -999 999 if you don't want
to winsorize or don't understand question>: -999 999
Want zero-start (say Z), steady state (say S) or FIR (say F)? : f
Enter ARL : 240
h  18.5000 arls  257.7   223.2   243.3
h  18.3000 arls  269.9   233.7   255.2
h  18.9000 arls  274.0   237.8   259.2
k  5.3000 h 18.9000 ARL  237.75
      h 19.0000 ARL  241.10
DI 19.000. IC ARL  241.1. OOC ARL  45.9. FIR ARL  30.8. SS ARL  33.6

```

Figura 7. 12 – Saída do programa anygeth para ARL com FIR.
 FONTE: Programa anygeth (HAWKINS e OLWELL, 1998).

Os resultados do trabalho de Morais e Pacheco (2006) reforçam que para um mesmo valor do limite superior ($h = 6$) um CUSUM individual tem $ARL_0 = 1015,71$ e o CUSUM com RIR, $ARL_0 = 995,07$. Assim, uma análise com o uso de RIR volta a ser citada no Capítulo 9, nas sugestões para trabalhos futuros.

¹⁸Disponível em <http://www.stat.umn.edu/users/cusum/software.htm>.



```

R Console
Arquivo  Editar  Misc  Pacotes  Janelas  Ajuda

> arlCusum(k=5.3,h=18.1,distr=c("binomial"),n=100,theta=0.05)
SARL
[1] 242.1569

SFIR.ARL
[1] 209.2863

> arlCusum(k=5.3,h=19,distr=c("binomial"),n=100,theta=0.05)
SARL
[1] 278.1923

SFIR.ARL
[1] 241.0979

```

Figura 7. 13 – Saída do R para ARL sem e com FIR.

7.5 ANÁLISE DOS RESULTADOS

Observando todos os gráficos em conjunto (Figuras 7.1 a 7.10), não se pode dizer que um deles é melhor que o outro em todas as situações. Na maioria dos casos, nenhum gráfico especificamente terá uma performance superior em comparação a outros gráficos para todas as magnitudes de mudança (REYNOLDS e STOUMBOS, 2006).

Os gráficos combinados são efetivos para aumentar a sensibilidade de esquemas CUSUM, para detectar alterações maiores. Isto significa que este gráfico tem bom desempenho para uma faixa maior de magnitudes de mudança. A parte CUSUM sinaliza mudanças próximas da planejada, neste trabalho fixadas como, de pequena magnitude (inferior a $s < 50\%$ ou $< 1\sigma$). O combinado tem desempenho igual (ou levemente inferior) aos esquemas CUSUM nas proximidades do valor de p_1 de projeto. As linhas de Shewhart permitem sinalizar mudanças maiores que as projeto. Existe uma região onde o gráfico combinado tem desempenho superior aos dois procedimentos individuais. É uma região com alterações moderadas, se comparadas com a de projeto, mas não grande o suficiente para ser detectada pelas linhas de Shewhart.

Se, contudo, a finalidade é sinalizar apenas alterações específicas de maior magnitude, como por exemplo, 200% em p_0 ($p_1 = 3p_0$) deve-se considerar um CUSUM planejado para tal, ou um gráfico do tipo Shewhart. Ressalta-se que há opções de esquemas CUSUM modificados especificamente para detecção de grandes alterações (WU; JIAO; LIU, 2008), sendo estes mais efetivos que o procedimento CUSUM convencional, para grandes alterações.

7.6 SÍNTESE DO CAPÍTULO

Neste capítulo o gráfico combinado foi comparado, em termos de ARL, com outros procedimentos CUSUM planejados para detectar alterações maiores. Foram realizadas duas análises, em que o gráfico combinado foi comparado com procedimentos CUSUM individuais e também, numa delas, com o procedimento Shewhart. O combinado aumenta a sensibilidade de um CUSUM para alterações maiores que as de planejamento. Se o processo requer apenas sinalizar rapidamente uma alteração específica, recomendam-se outros procedimentos, como por exemplo, o CUSUM individual.

8. PLANEJAMENTO DE UM GRÁFICO COMBINADO E APLICAÇÕES

Este capítulo traz algumas aplicações de um gráfico combinado Shewhart-CUSUM. Inicialmente, são sugeridos alguns procedimentos para o planejamento de um gráfico combinado Shewhart-CUSUM para atributos. Esta seqüência de passos leva em conta as suposições necessárias para a aplicação do gráfico. Em seguida estão três aplicações. A primeira foi adaptada de um exemplo da literatura; a segunda ao monitoramento de um indicador da qualidade de água potável e a terceira envolve o controle de qualidade no processo de rotulagem de garrafas. Para complementar, os gráficos combinados são comparados com procedimentos Shewhart e CUSUM individuais.

8.1 PASSOS PARA CONTRUÇÃO DO GRÁFICO COMBINADO

Nesta seção estão alguns procedimentos sugeridos para a construção de um gráfico combinado Shewhart-CUSUM binomial. Estes passos foram adotados para as aplicações do gráfico deste capítulo. Seja uma série de N amostras de tamanho n com proporção sob controle p_0 . Para a construção do gráfico deve-se:

- 1) definir qual a proporção p_1 que se deseja detectar;
- 2) determinar ARL_0 mínimo tolerável (ou taxa de alarmes falsos máxima) para o gráfico combinado;
- 3) fazer as seguintes verificações:
 - i) aderência à distribuição binomial;
 - ii) ausência de superdispersão;
 - iii) independência / ausência de autocorrelação;
- 4) calcular os valores de ARL_0 para os gráficos individuais – Equação(4.5);
- 5) determinar o valor referência k – Equação (3.11);
- 6) aproximar o valor do limite de decisão h com o ARL_0 para a parte CUSUM pela aproximação CD;
- 7) calcular o LSC do gráfico “np” através de:
 - i) limites convencionais quando possível;
 - ii) limites probabilísticos (exatos);

8) verificar se $LSC \leq h + k$, caso contrário, não é um gráfico combinado;

Um fluxograma dos passos está em sequência (Figura 8.1).

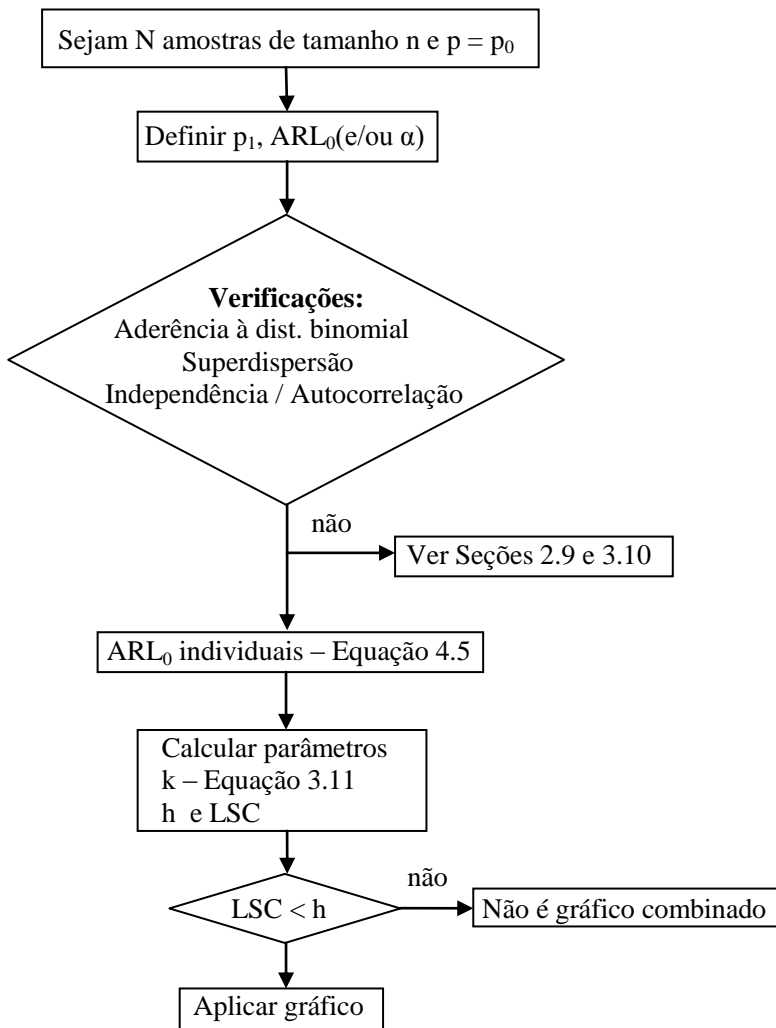


Figura 8. 1 – Fluxograma para construção de um gráfico combinado.

Nesta sequência de passos foi definida a abordagem para a aproximação do limite superior mediante a aproximação CD (ver Capítulo 5), que é a utilizada neste Capítulo. Ressalta-se que outras abordagens, como por exemplo, cadeias de Markov, podem ser aplicadas.

Duas rotinas no R foram escritas para construção de um gráfico combinado Shewhart-CUSUM binomial, unilateral superior e um CUSUM binomial unilateral superior. Estas rotinas são aplicadas nos exemplos e estão nos Apêndice F e G, respectivamente. Os códigos no R para as aplicações estão no Apêndice H.

8.1.1 Suposições

Quando não for verificada a aderência à distribuição binomial, não se aconselha a aplicação do gráfico. A parte CUSUM é muito sensível ao afastamento da distribuição (HAWKINS e OLWELL, 1998). Se houver superdispersão, que afeta ambas as partes, CUSUM e Shewhart, recomenda-se sua remoção. Isto pode ser feito mediante a aplicação do gráfico aos resíduos de um modelo linear generalizado, como por exemplo, um modelo binomial. Mas até o momento não se tem registros de trabalhos abrangendo gráficos combinados nesta área.

Em se tratando de dependência, em sequências no tempo, como é caso da maioria das aplicações em gráficos de controle, deve-se observar a ausência de autocorrelação nos dados. Existindo autocorrelação, gráficos que a captem, para as partes individuais, são alternativas. Embora existam diversos trabalhos para gráficos Shewhart e CUSUM para dados discretos autocorrelacionados, não há ainda estudos com gráficos combinados. Nas Seções 2.9 e 3.10 estão mais detalhes sobre este assunto, que fornecem opções aos procedimentos individuais para dados autocorrelacionados.

8.1.2 Outras distribuições

Embora os procedimentos aqui descritos sejam voltados para variáveis com distribuição binomial, eles podem ser aplicados a dados com distribuição de Poisson, com poucas modificações. No passo 1, ao invés da proporção p_1 , deve se definir λ_1 (média) fora de controle estatístico que se deseja detectar. As verificações são as mesmas e os

parâmetros k e h têm expressões e abordagens específicas. O limite LSC pode ser convencional ou probabilístico levando em conta a distribuição de Poisson.

8.2 APLICAÇÃO 8.1: XÍCARAS DEFEITUOSAS

Nesta seção, um gráfico combinado é aplicado aos dados adaptados de um problema de Devore (2006). O conjunto de dados se compõe de uma amostra de $n = 100$ xícaras de chá de um determinado jogo de jantar, selecionadas a cada 25 dias sucessivos. A quantidade de xícaras defeituosas encontradas (por dia) está na Tabela 8.1.

Tabela 8.1 - Número de xícaras defeituosas por amostra.

Dia	1	2	3	4	5	6	7	8	9	10	11	12	
x	7	4	3	6	4	9	6	7	5	3	7	8	
Dia	13	14	15	16	17	18	19	20	21	22	23	24	25
x	4	6	2	9	7	6	7	11	6	7	4	8	6

Fonte: Devore (2006:629).

Estes dados são usados para a fase I, quando são definidos os limites de controle. A proporção sob controle p_0 é 0,0608 e para p_1 considerou-se $p_1 = 0,073 (s = 20\%, 0,51\sigma)$. O valor de $ARL_0 = 370$ ($\alpha = 0,27\%$). Assim um gráfico combinado foi planejado com $k = 6,7$, $h = 17,1 (ARL_0 = 425)$ e $LSC = 15 (ARL_0 \cong 2852)$. Como estes dados provêm da literatura, não foram verificadas as suposições.

Para a fase II, que é o monitoramento, é simulada uma situação fora de controle estatístico, com $p_1 = 0,073$ e uma alteração propositalmente grande ($s \cong 180\%$) na 30ª amostra (5ª amostra após a mudança em p_0). Garantiu-se que a média dos dados simulados fosse igual a np_1 . Aplicou-se o gráfico combinado aos dados e o resultado está na Figura 8.2.

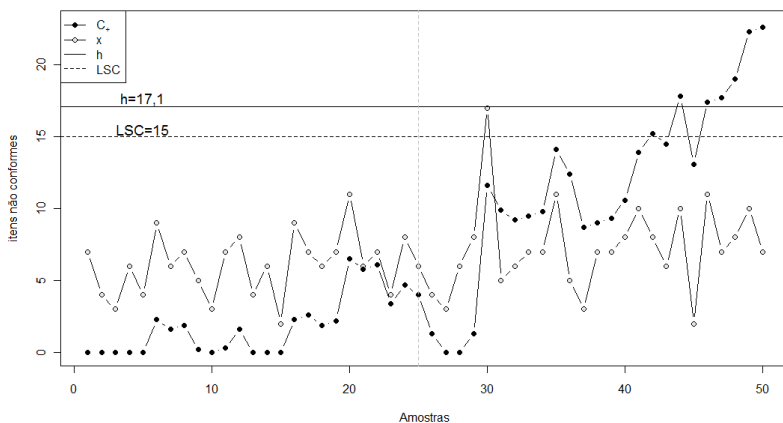


Figura 8. 2 – Gráfico combinado para os dados de xícaras defeituosas.

O gráfico combinado sinaliza na parte Shewhart a alteração na 30ª amostra e a parte CUSUM evidencia o aumento na média do processo (para $np_1 = 7,3$) com h ultrapassado na 44ª amostra.

Em seguida os resultados do gráfico combinado são comparados com gráficos CUSUM e Shewhart individuais. O gráfico CUSUM tem parâmetros $k = 6,7$ e $h = 16,5$ ($ARL_0 \cong 370$). O valor de LSC do gráfico Shewhart é $LSC = 13,25$ (DEVORE, 2006). O gráfico CUSUM está na Figura 8.3 e o gráfico Shewhart na Figura 8.4¹⁹.

O gráfico CUSUM, Figura 8.3, sinaliza o aumento da média np_0 , com o limite h ultrapassado na 44ª amostra, como no combinado. O CUSUM, no entanto, não sinaliza a alteração repentina na 30ª amostra. O gráfico Shewhart, Figura 8.4, sinaliza apenas a alteração da 30ª amostra, mas não consegue detectar o aumento da média, nem com as regras sensibilizantes. Este é um exemplo em que se mostra que o gráfico combinado é mais efetivo que os gráficos individuais.

¹⁹Os gráficos das Figuras 8.4, 8.9 e 8.16 foram gerados com o pacote qcc (SCRUCCA, 2004) do aplicativo GNU R, que inclui as regras sensibilizantes (Seção 2.7). Para os demais foram feitas rotinas no R (Apêndice G).

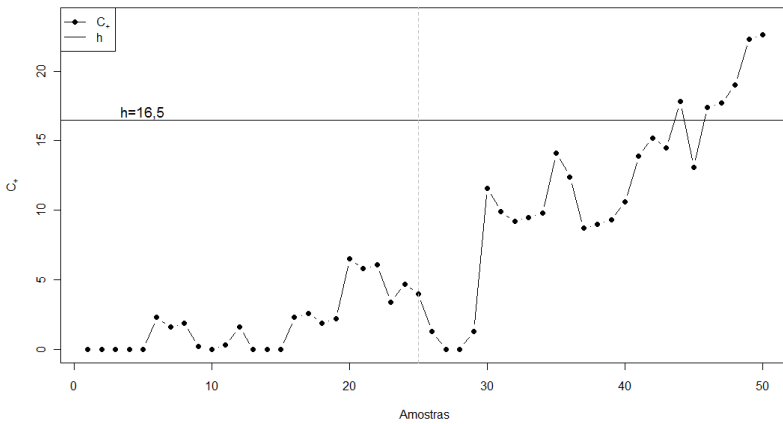


Figura 8. 3 – Gráfico CUSUM para os dados de xícaras defeituosas.

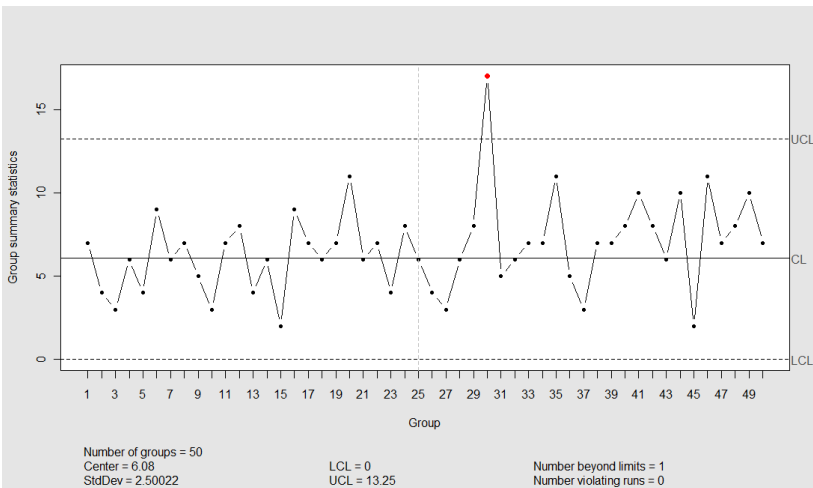


Figura 8. 4 – Gráfico tipo Shewhart para os dados de xícaras defeituosas.

8.2 APLICAÇÃO 8.2: INDICADOR DE COLIFORMES TOTAIS

A aplicação a seguir abrange a análise do monitoramento de Coliformes Totais, na rede de distribuição de água potável da Companhia Águas de Joinville. Este parâmetro (coliformes totais) é um indicador da possibilidade da existência de microorganismos patogênicos (causadores de doenças). De acordo com a Portaria N^o 518/2004, do Ministério da Saúde (BRASIL, 2004) é tolerada a presença de coliformes totais em até 5% das amostras analisadas.

As duas estações de tratamento de água (ETA) de Joinville têm um rigoroso controle, sendo analisadas amostras de controle operacional e de qualidade nas águas brutas (água do rio), decantada e filtrada (fases intermediárias) e tratadas. Saindo da ETA, a água de abastecimento vai para a rede de distribuição.

As análises realizadas na rede de distribuição de água devem estar de acordo com a Portaria N^o 518/2004, do Ministério da Saúde (Brasil, 2004). O plano de amostragem também é definido pela portaria, que exige um número mínimo de análises mensais a serem realizadas em função da população abastecida. No caso de Joinville são exigidas 200 análises para diversos parâmetros, entre eles, o de coliformes totais.

São 205 pontos de coleta, situados em pontos diversos da cidade. As coletas são realizadas em dias e horários pré-definidos, sendo a amostras classificadas como conformes (não apresentam coliformes totais) e não conformes (apresentam coliformes totais). Ao final conta-se o número de amostras não conformes. Os relatórios de análise são mensais. A empresa, embora tenha uma metodologia de controle de qualidade, não aplica nenhum gráfico de controle estatístico de processo.

A série analisada corresponde ao período de 2007, 2008 e 2009, com frequência mensal. O conjunto de dados está no Anexo F. Os dois primeiros anos, 2007 e 2008, correspondem à fase I do processo, definida no Capítulo 2. O valor de p , considerado sob controle, é $p_0 = 0,022$. O monitoramento (segunda fase do processo) será feito quando analisadas as observações do ano de 2009.

8.2.1 Verificação das Suposições

Para aplicar os gráficos combinado, CUSUM e Shewhart é necessário verificar as suposições de independência, aderência e superdispersão. Efetuando um teste formal, baseado na estatística qui quadrado, verifica-se a ausência de superdispersão ($p\text{-valor} = 0,2850$)²⁰. A aderência à distribuição binomial foi verificada também mediante dois testes estatísticos formais: um embasado na razão de verossimilhança ($p\text{-valor} = 0,752$)²¹ e o teste qui quadrado ($p\text{-valor} = 0,757$).

A independência pode ser avaliada pelo processo de amostragem, uma vez que as coletas são realizadas em locais, dias e horários distintos, portanto independentes. O teste de Wald-Wolfowitz (de aleatoriedade) foi aplicado e reforça a suposição de independência dos dados ($p\text{-valor} = 0,2122$). Foram também construídos os gráficos da função de autocorrelação e autocorrelação parcial amostral. Os gráficos, Figura 8.5, mostram que os dados não possuem autocorrelação. Portanto, os gráficos Shewhart, CUSUM e combinado podem ser aplicados aos dados.

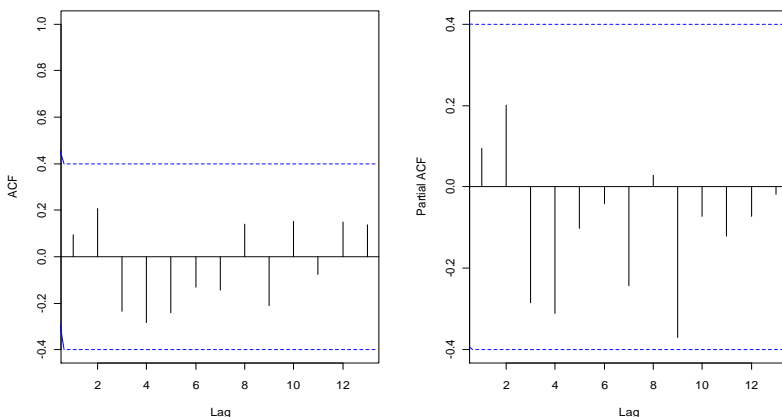


Figura 8. 5 – Gráficos da função de autocorrelação e autocorrelação parcial para os dados de coliformes totais.

²⁰Os testes de aderência estão no Apêndice G (rotinas) e I (resultados).

²¹O nível de significância adotado em todos os testes foi de 5%.

8.2.2 Construção e Aplicação do Gráfico Combinado

Para o planejamento dos gráficos, inicialmente pensou-se em considerar um $ARL_0 = 370$ ($\alpha = 0,27\%$), mas não foi possível atingir este valor no gráfico do tipo Shewhart. O limite convencional de Shewhart é 10,65 e não alcança o $ARL_0 = 370$ ($ARL_0 \cong 163$). Aplicou-se o limite exato (ou probabilístico) $LSC = 11$, $ARL_0 = 465,5$, mas nenhum dos dois limites não chega próximo ao ARL_0 de projeto. Escolheu-se então para este trabalho um ARL_0 de 465,5 ($\alpha = 0,21\%$).

Para a construção do gráfico combinado foi necessário definir a proporção p_1 que se deseja detectar. Definiu-se $s = 20\%$ e $p_1 = 0,026$ ($0,6\sigma$). Assim, acredita-se que o combinado irá detectar alguma alteração, de pequena a moderada na média (np) do processo e também será sensível a grandes alterações, no caso $p = 0,05$ ($s = 127\%$, 3σ) que é o limite estabelecido pela legislação. A parte CUSUM do combinado terá $ARL_0 = 469$ ($\alpha = 0,22\%$) e a parte Shewhart $ARL_0 = 65817$ ($\alpha \cong 0,00000152$). Os parâmetros da parte CUSUM são $k = 4,95$ e $h = 17,6$. O limite superior do gráfico Shewhart é $LSC = 15$ (probabilístico). Verifica-se também a condição $LSC < h + k$. Os valores de ARL_0 recalculados com os parâmetros são iguais a 473 ($\alpha = 0,21\%$), 65817 ($\alpha \cong 0,0000152$) e 469 ($\alpha = 0,22\%$), para as partes CUSUM, Shewhart e o procedimento combinado.

O gráfico resultante está na Figura 8.6. As primeiras 24 amostras correspondem ao período considerado sob controle estatístico (anos de 2007 e 2008) e as 12 restantes, ao período que se deseja monitorar (2009). A parte CUSUM do combinado indica que a qualidade do processo está se deteriorando. Há evidências de aumento do número de amostras não conformes, ao que sugere que a média (np) do processo mudou. A linha superior de controle é ultrapassada na 35ª amostra (mês de julho de 2009), tanto na parte CUSUM como na parte Shewhart. Porém, ao se observar apenas a parte CUSUM, a tendência de crescimento aparece a partir da 27ª amostra (março de 2009), não retornando mais à condição de controle estatístico.

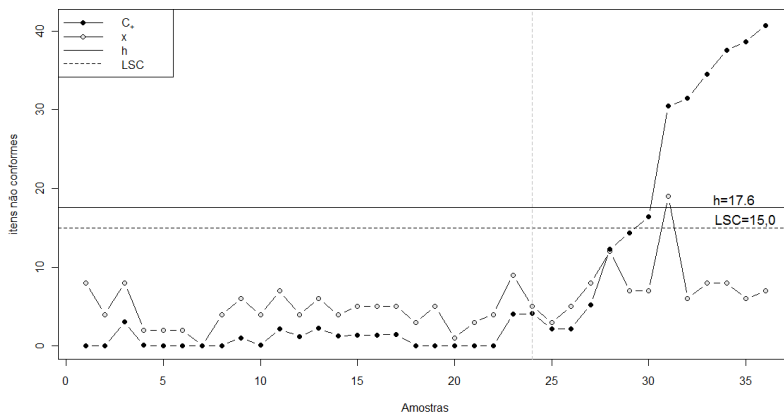


Figura 8. 6 – Gráfico Combinado Shewhart-CUSUM para os dados de Coliformes Totais.

8.2.3 Comparação com outros gráficos

Para fins de comparação, foi construído um gráfico Shewhart individual com o mesmo $ARL_0 = 465,5$ e $LSC = 11,0$. O gráfico do tipo Shewhart, Figura 8.7, sinaliza antes do combinado (28ª amostra – abril de 2009), uma mudança na média (np) sob controle.

Um gráfico CUSUM planejado para detectar $p_1 = 0,05$ ($s = 127\%$) que é o limite estabelecido pela resolução, com $k = 6,9$, $h = 5,21$ e $ARL_0 = 465,5$ está na Figura 8.8. Este gráfico tem o mesmo desempenho que o Shewhart, sinalizando, na 28ª amostra (abril de 2009), que indica, ainda, uma tendência de aumento das análises não conformes, o que não aparece no gráfico do tipo Shewhart. Pode-se concluir então que a média (np) do processo mudou e aumentou.

Ao serem aplicadas as regras suplementares ao procedimento de Shewhart, elas também indicam que o processo está fora de controle estatístico, amostras 32 a 36 (ver Figura 8.9). O uso destas regras, no entanto, implica aumento do risco α , ou seja, há uma probabilidade maior de ocorrência de alarmes falsos. Além disso, a interpretação do gráfico se torna mais complexa.

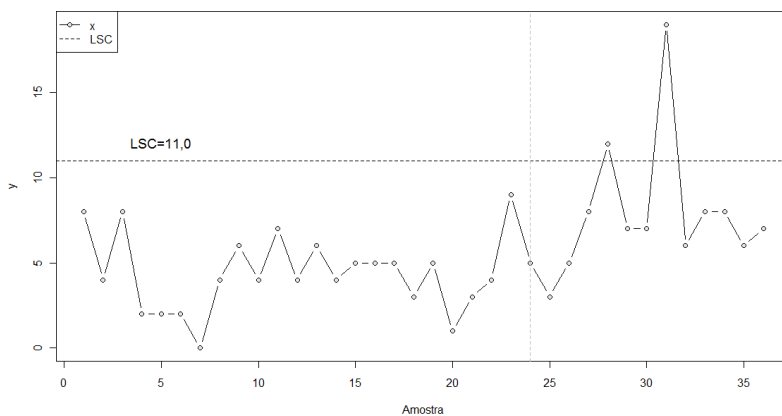


Figura 8. 7 – Gráfico Shewhart para os dados de Coliformes Totais.

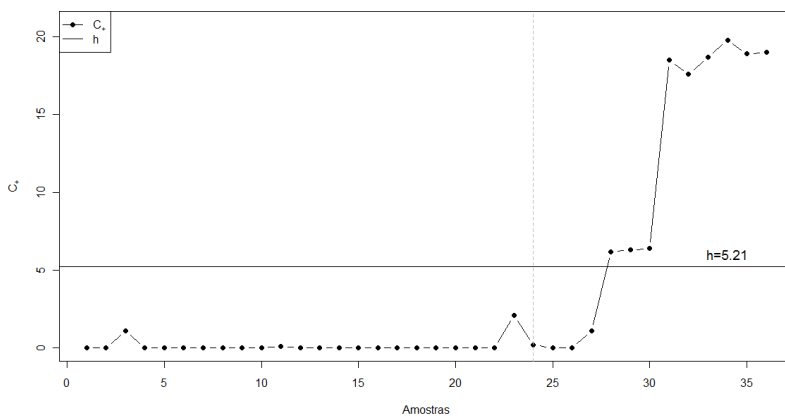


Figura 8. 8 – Gráfico CUSUM para os dados de Coliformes Totais.

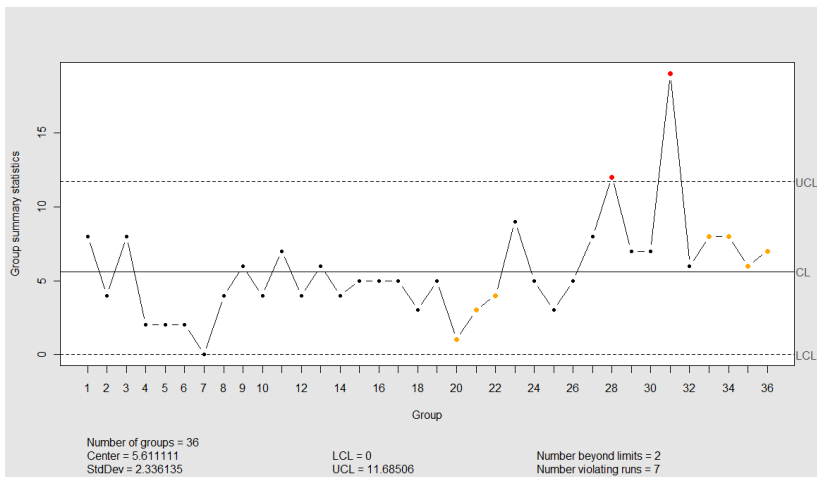


Figura 8. 9 – Gráfico Shewhart para os dados de Coliformes Totais com regras suplementares.

Para complementar o estudo foi planejado um gráfico combinado em que a parte CUSUM foi ajustada para um ARL_0 maior (519, $\alpha = 0,19\%$) e a parte Shewhart para um ARL_0 menor (4809, $\alpha = 0,021\%$) (Figura 8.10).

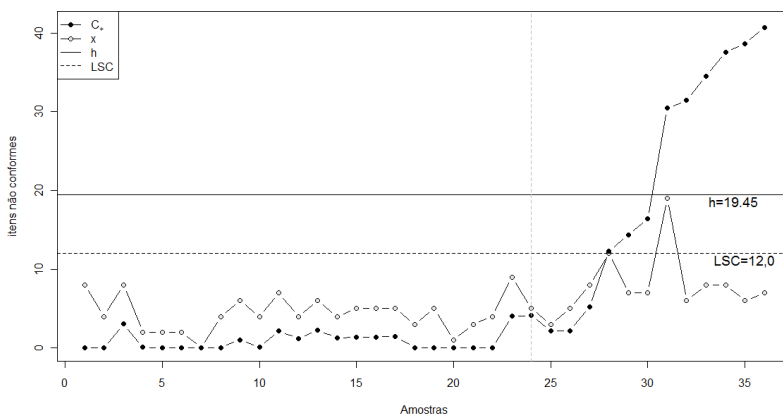


Figura 8. 10 – Gráfico Combinado Shewhart-CUSUM com $ARL_0 = 465,5$ e limites ajustados.

Nota-se que no gráfico da Figura 8.10, não houve mudanças significativas em relação ao gráfico da Figura 8.6. Mesmo com o ajuste dos limites o combinado sinaliza apenas na 7ª amostra (julho de 2009).

Ajustar novamente os limites, ou seja, fazer $LSC=11$, ou menos implicaria um ARL_0 muito alto para o CUSUM, superior a 20000. Tanto o gráfico Shewhart como o procedimento CUSUM individualmente seriam mais efetivos.

Todos os procedimentos, isto é, combinado, CUSUM e Shewhart, indicaram que a média (μ) do processo mudou, o que não é desejável. Implica que coliformes totais estão presentes em um número maior de análises e, por isso, são recomendáveis investigações na linha de distribuição.

Nesse exemplo de aplicação, foi possível observar outros aspectos interessantes:

i) O gráfico combinado não teve o melhor desempenho, sendo superado pelo Shewhart individual e pelo CUSUM individual, mesmo quando se balanceavam os limites;

ii) Há dificuldade em se estabelecer os limites do gráfico Shewhart, para se alcançar o ARL_0 de projeto;

iii) O desempenho do CUSUM individual se igualou ao gráfico Shewhart, com a vantagem de detectar a tendência de aumento no processo. Neste último, isto só é possível com auxílio das regras suplementares.

Assim, um gráfico CUSUM planejado para detecção de p_1 correspondente a 0,05 (5%) pode ser recomendado para o monitoramento deste particular processo. Nas análises realizadas foi tão efetivo quanto o gráfico do tipo Shewhart, mas com a vantagem de trazer informações sobre a tendência do processo.

8.3 APLICAÇÃO 8.3: PROCESSO DE ROTULAGEM DE GARRAFAS

O terceiro exemplo corresponde à aplicação do gráfico combinado a um processo de rotulagem de garrafas de uma marca de bebida destilada, pertencente a uma empresa do ramo, de grande porte. Os dados correspondem às unidades não conformes resultantes no

processo de colocação dos rótulos e contra-rótulos nas garrafas, após o enchimento.

O rótulo representa o conjunto de designações, ilustrações ou marcas que caracterizam a bebida. No rótulo devem estar informações obrigatórias como marca, teor alcoólico, volume, aditivos, prazo de validade, composição e registros. O contra-rótulo é considerado uma extensão do rótulo (RIZZON e MENEGUZZO, 2006).

Quando o rótulo está rasgado, descentrado, amassado, descolando ou com falha na impressão, considera-se o produto defeituoso, devendo a garrafa ser lavada e retornar ao processo de rotulagem, o que era classificado como retrabalho.

Na época em que os dados foram coletados, a empresa iniciava um processo de Controle Estatístico de Qualidade, com a aplicação de Gráficos de Shewhart com limites convencionais. Além do processo de rotulagem, outras etapas na fabricação desta e de outras bebidas foram incluídas nos trabalhos.

Para a aplicação do gráfico foram utilizadas séries dados de quatro dias consecutivos, sendo o primeiro para a fase I, e os três seguintes para a fase II. Não corresponde exatamente ao que foi realizado na empresa na época, mas servirá ao propósito deste trabalho. As amostras eram sequenciais, e uma caixa com 36 unidades era retirada de cada lote, totalizando aproximadamente 18 amostras por dia. Se houvesse indicação de estar fora de controle estatístico, o lote retornava, para avaliação e posterior lavagem e nova rotulação. O método de inspeção é visual. O conjunto de dados está no Anexo G.

8.3.1 Verificação das suposições

Como as amostras eram sequenciais, no tempo, e pertenciam à mesma caixa, o processo de amostragem não dava nenhuma garantia de aleatoriedade. A construção dos gráficos da função de autocorrelação e autocorrelação parcial amostral, Figura 8.11, indica a ausência de autocorrelação no processo. Para complementar o teste de Wald-Wolfowitz foi realizado para reforçar a hipótese de independência (p -valor = 0,8454). Também foi verificada a ausência de superdispersão (p -valor = 0,4044). A aderência à distribuição binomial foi avaliada mediante dois testes: quiquadrado (p -valor = 0,139) e baseado na razão de verossimilhança (p -valor = 0,116).

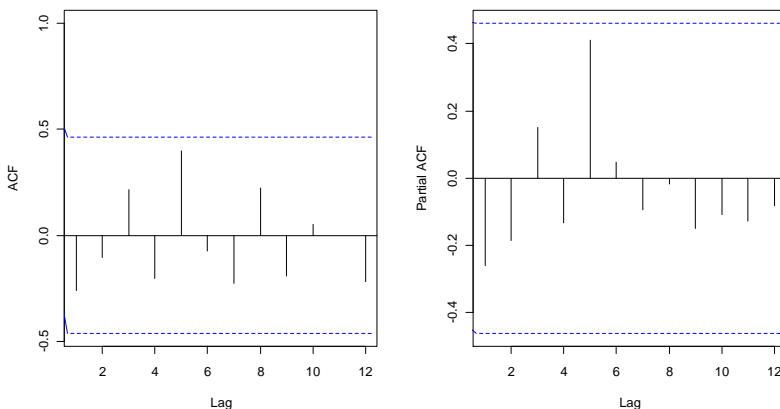


Figura 8. 11 – Gráficos da função de autocorrelação e autocorrelação parcial para os dados da rotulagem de garrafas.

8.3.2 Aplicação do Gráfico Combinado

Conforme informação anterior a empresa aplicava gráficos tipo Shewhart com limites 3σ convencionais ($ARL_0 = 370$, $\alpha = 0,27\%$), e $LSC = 5,9$. O tamanho da amostra é $n = 36$ e a proporção considerada sob controle é $p_0 = 0,052$. Para estes dados o valor de ARL_0 alvo não é alcançado $P(x \geq LSC(5,9)) = 0,0023$. Para fins de comparação, do combinado com os outros gráficos, inclusive o de Shewhart, o ARL_0 alvo adotado será de 443 ($\alpha = 0,23\%$). Assim, o limite convencional ($LSC = 5,9$) é mantido e está muito próximo do limite probabilístico ($LSC = 6$).

A probabilidade p_1 , fora de controle estatístico, necessária para a parte CUSUM é de $p_1 = 0,07$ ($s \cong 37\%$; $0,48\sigma$). Para a definição deste valor considerou-se detectar uma mudança de cerca de meio desvio padrão na média “np”. Optou-se também por um combinado com os limites de Shewhart não muito “largos”. O gráfico será menos sensível a pequenas variações, mas acredita-se que irá captar mudanças moderadas a grandes. A intenção é não relaxar muito o limite superior LSC de. Assim a parte CUSUM tem $ARL_0 = 548$, com $k = 2,18$ e $h = 11,3$. A parte Shewhart tem $ARL_0 = 2298$, com $LSC = 7$. Recalculando

(mediante cadeias de Markov) os valores de $ARL_0(\alpha)$ são 447,8 (0,22%) para o combinado e 556 (0,18%) para a parte CUSUM.

Na Figura 8.12 está o gráfico combinado Shewhart-CUSUM. A linha tracejada vertical (18ª amostra) separa os períodos sob controle e o de monitoramento. A parte Shewhart do gráfico combinado emite um sinal na 27ª e depois na 41ª amostras. Avaliando os dados, nestas duas situações, constata-se que $\frac{1}{4}$ dos itens analisados são defeituosos. A parte CUSUM dá informações sobre a tendência do processo, indicando que há um aumento na média “np” de itens conformes. A partir da 27ª amostra o CUSUM não retorna ao zero, e verifica-se aos poucos uma tendência crescente. Nas amostras 41 e 45 a linha do CUSUM é ultrapassada, sendo que após esta última amostra o processo não retorna ao estado de controle. Há evidências de que a média “np” do processo aumentou, o que resulta no deterioramento da qualidade do processo de rotulagem. Isso implica retrabalho para a empresa.

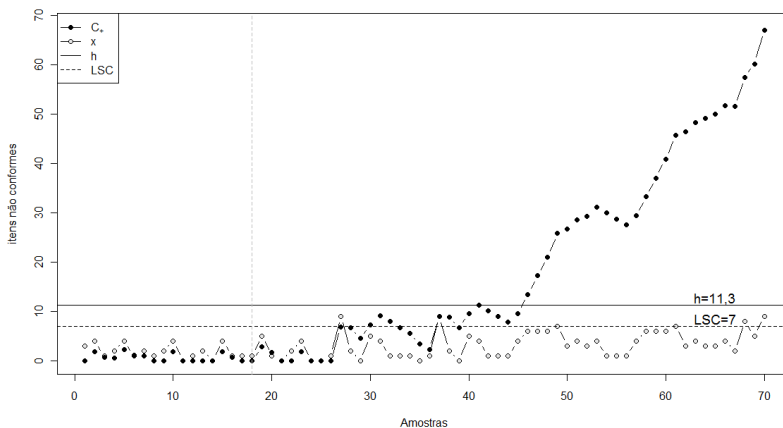


Figura 8. 12– Gráfico Shewhart-CUSUM para os dados da rotulagem de garrafas.

8.3.3 Comparação com outros gráficos

Para os mesmos dados e situação, $p_1 = 0,07$ ($s = 35\%$) e $ARL_0 = 443$, foi planejado e construído um gráfico CUSUM. O valor de referência é o mesmo ($k = 2,18$), e o limite superior é $h = 10,6$. O ARL_0 recalculado (Markov) é 442,9. O

gráfico está na Figura 8.13 e nele pode-se verificar que praticamente há as mesmas informações da parte CUSUM do combinado.

Um segundo gráfico CUSUM individual foi feito. Este tem um valor p_1 de projeto maior ($s = 100\%$, $p_1 = 2p_0$), mas com o mesmo ARL_0 (Figura 8.14).

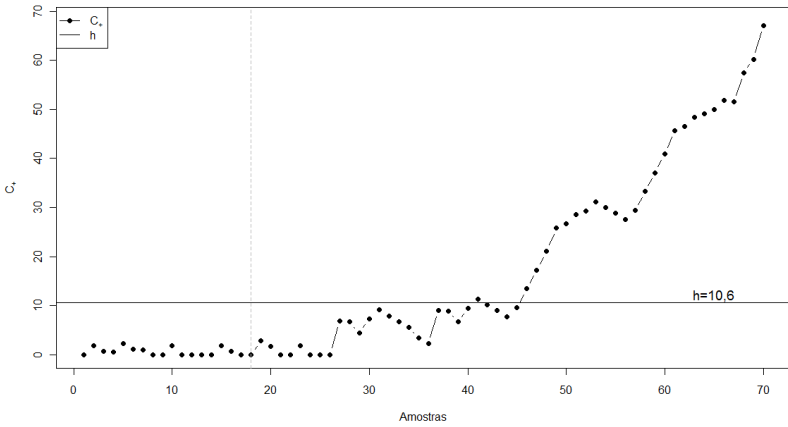


Figura 8. 13 – Gráfico CUSUM para os dados da rotulagem de garrafas.

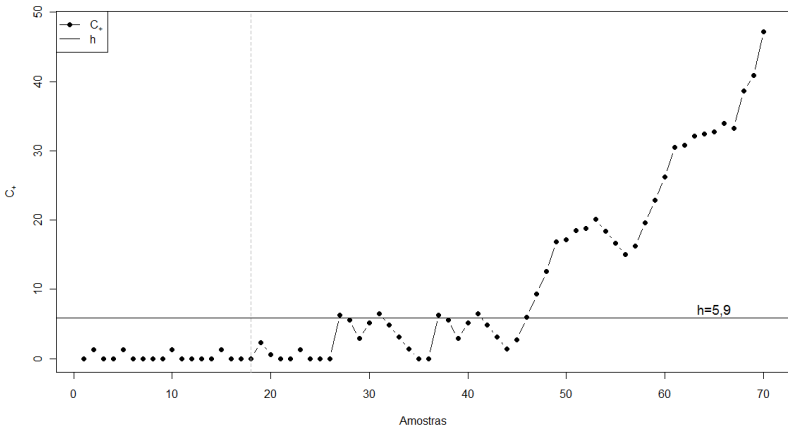


Figura 8. 14 – Gráfico CUSUM para os dados da rotulagem de garrafas

O valor de referência é $k = 2,7$ e o limite superior é $h = 5,9$, retornando $ARL_0 = 455$. No gráfico, Figura 8.14, visualiza-se que o limite superior é ultrapassado na amostra 27. Depois a tendência de aumento é evidenciada pela linha do CUSUM.

Em seguida é analisado o gráfico Shewhart, com $LSC = 6$. Pode ser visto que, considerando apenas o limite superior, somente as amostras com mais itens defeituosos são sinalizadas (Figura 8.15). Não há indicação do aumento da média do processo.

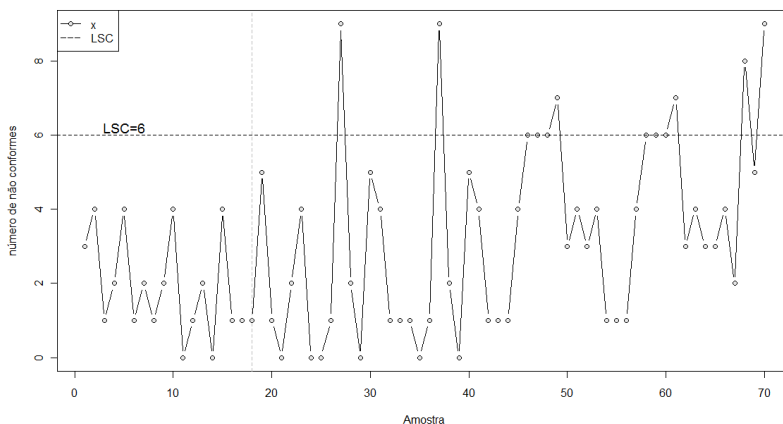


Figura 8. 15 – Gráfico Shewhart para os dados da rotulagem de garrafas.

Na Figura 8.16 está um gráfico Shewhart com as regras sensibilizantes. Estas dão suporte à idéia de que algo no processo está ocorrendo, mas não são diretamente explícitas quanto à tendência.

O gráfico combinado, para esta aplicação é uma opção interessante. Observa-se que ele teve o mesmo desempenho que o gráfico do tipo Shewhart, na sinalização de alterações maiores e repentinas (amostras 27 e 41). E, comparado ao CUSUM, permitiu como este, detectar que a média “np” do processo mudou.

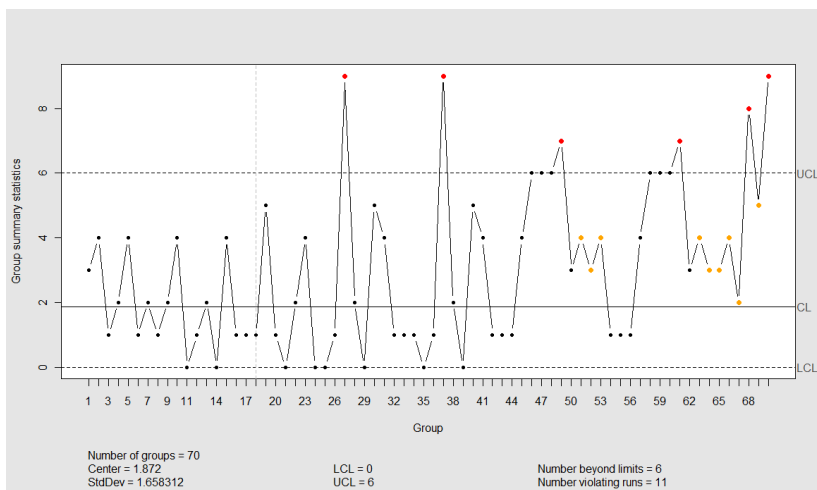


Figura 8. 16 – Gráfico Shewhart com regras sensibilizantes para os dados da rotulagem de garrafas.

O gráfico CUSUM para mudanças maiores, como mostra a Figura 8.14, teve a mesma performance que o combinado. Mas não se pode dizer que isto vale para mais situações. Precisa ser analisado o desempenho na sinalização de diversas magnitudes de mudança.

8.4 SÍNTESE DO CAPÍTULO

Neste capítulo o gráfico combinado Shewhart-CUSUM foi aplicado para o monitoramento de três conjuntos de dados: um adaptado da literatura e dois casos de dados reais. Os resultados foram comparados com a aplicação de procedimentos Shewhart e CUSUM individuais. As particularidades de cada caso levam a conclusões diferentes. No primeiro estudo de caso foi criada uma situação fora de controle estatístico envolvendo uma mudança grande e repentina e outra pequena e gradativa. A parte Shewhart sinalizou a primeira e a parte CUSUM do combinado a segunda, ou seja, o combinado sinalizou as duas. Comparando tal gráfico com os procedimentos individuais, verificou-se que o gráfico Shewhart sinalizou apenas a alteração repentina e o CUSUM o aumento na média do processo. No segundo caso, de monitoramento de coliformes totais, o gráfico combinado não

apresentou desempenho melhor que os procedimentos individuais. Na terceira aplicação, no processo de rotulagem, o gráfico combinado se mostrou uma alternativa interessante. No processo de rotulagem, o gráfico combinado captou alterações repentinas e persistentes. O gráfico Shewhart, sem as regras suplementares, sinalizou apenas as mudanças de maior magnitude e repentinas. O gráfico CUSUM detectou a mudança no nível da média (\bar{np}) do processo.

O gráfico combinado, ainda, foi comparado com o gráfico CUSUM planejado para alterações maiores e Shewhart com regras sensibilizantes. Para ambas as aplicações, estes últimos gráficos tiveram o mesmo comportamento do combinado. No entanto, deve-se levar em conta as especificidades de cada processo. Um gráfico CUSUM para mudanças maiores pode não responder a pequenas alterações e o uso de regras suplementares tem as suas deficiências, já abordadas neste e em outros Capítulos.

A escolha de um gráfico para o controle e monitoramento de um processo depende das características desse processo. Por isso, é fundamental que se conheça o processo, o que é necessário controlar e o que se espera com o monitoramento.

9. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Este trabalho tratou de gráficos combinados Shewhart-CUSUM para atributos, especificamente gráficos do tipo “np” em que a distribuição subjacente é a binomial. O objetivo principal era contribuir para o desenvolvimento e aperfeiçoamento de um gráfico combinado Shewhart-CUSUM binomial.

Os objetivos secundários abrangiam: o desenvolvimento de uma aproximação do limite superior h do CUSUM binomial, desenvolver uma metodologia para encontrar as medidas de desempenho do gráfico de controle combinado unilateral superior, investigar o comportamento do RL do gráfico combinado, verificar se, quando aplicado, o gráfico combinado atende aos parâmetros de projeto, avaliação do desempenho do gráfico proposto comparando-os com procedimentos individuais e aplicação do gráfico combinado a dados reais.

No Capítulo 2 foi feita uma revisão de gráficos de controle, especificamente os gráficos de Shewhart para atributos. Chamou-se a atenção para o fato de muitas vezes os limites convencionais não atingem o valor de ARL_0 . Como alternativa estão os limites exatos (ou probabilísticos) que foram aplicados quando necessário na construção do gráfico combinado. As suposições de aderência, superdispersão e, principalmente independência, foram abordadas incluindo sugestões de testes formais e métodos de reconhecimento sugeridos pela literatura.

A principal contribuição do Capítulo 3 foi a de analisar o gráfico CUSUM sob a ótica da família exponencial de distribuições. As propriedades da família exponencial tornam mais simples a compreensão de algumas características do procedimento, como por exemplo, o valor de referência k , a aproximação do ARL sugerida por Siegmund (1985), extensões dessa aproximação por Reynolds e Stoumbos (1999, 2001), e inclusive a própria abordagem markoviana do ARL, refinada por Hawkins (1992).

No Capítulo 4, a revisão bibliográfica sobre os gráficos combinados trouxe alguns importantes subsídios: uma formulação simples de calcular o ARL_0 do gráfico combinado balanceando os valores de ARL_0 dos esquemas individuais; a posterior aproximação dos parâmetros e a possibilidade de comparação de alguns resultados.

Uma das proposições iniciais, uma proposta para cálculo do limite superior h do CUSUM binomial, está no Capítulo 5. A aproximação com origem nos trabalhos de Siegmund (1985) e Reynolds

e Stoumbos (1998, 1999, 2001) para o ARL vista no Capítulo 3, é desenvolvida de modo a permitir o cálculo do limite superior h , com auxílio da função LambertW e seu desenvolvimento em série. Foi proposta uma Equação com quatro termos (no desenvolvimento) e avaliada a acurácia desta mediante simulação e comparação dos resultados com valores da literatura e com a abordagem markoviana. A Equação, embora seja uma alternativa às formas usuais de aproximação de h , não produz bons resultados para valores de referência k grandes e para valores muito pequenos de s (menores de 5%). Não é fácil de determinar o que é um valor de k grande, já que para a distribuição binomial o tamanho da amostra n , na formulação adotada, influencia no valor. Verificou-se, no entanto, que na maioria dos casos a aproximação não se distancia muito do valor de h , num intervalo de ± 1 (um). Elaborou-se um algoritmo em que este h aproximado serve de valor inicial para um processo iterativo englobando cadeias de Markov e interpolação logarítmica. De modo geral, no máximo duas iterações são suficientes.

No Capítulo 6 estudou-se o comportamento do RL do gráfico combinado e seu desempenho comparando-o com o CUSUM individual e com gráfico Shewhart, quando possível. Para analisar o RL do gráfico combinado, procedimentos de simulação foram aplicados e as partes individuais avaliadas. O RL de um gráfico combinado aproxima-se do RL da parte CUSUM para pequenas alterações e na medida em que a magnitude das alterações é maior incorpora o RL da parte Shewhart. De modo geral o RL é assimétrico, platicúrtico e com valores altos de SDRL.

Ainda, no Capítulo 6, com relação ao desempenho, também avaliado com simulações, verificou-se que a adição das linhas Shewhart realmente aumenta a sensibilidade de um CUSUM binomial para alterações maiores. Nas análises o ARL_0 da parte CUSUM tinha valor próximo ao do combinado, com parâmetro h próximo ao do CUSUM isolado. Os limites do gráfico “np” foram balanceados para manter o ARL_0 alvo (do combinado e CUSUM) previsto. São limites mais “abertos” que os dos gráficos “np” individuais. E, em grande parte das análises, foram utilizados limites exatos (ou probabilísticos). Quando possível (por causa do valor de ARL_0) comparou-se o gráfico combinado com um gráfico individual de Shewhart. Assim verificou-se que o gráfico combinado tem desempenho superior a ele para alterações menores (inferiores a 2σ em relação média np_0 nos casos analisados). Identificou-se também ir existir uma regio onde o grfico combinado

terá desempenho superior aos gráficos CUSUM e tipo Shewhart. Foi possível também constatar, por meio das simulações realizadas que o ARL_0 de planejamento foi alcançado. Para o ARL_1 , não há ainda como comparar os valores.

No Capítulo 7, o desempenho do gráfico combinado, em termos de ARL, foi comparado com procedimentos CUSUM (e Shewhart, quando possível) planejados para detectar mudanças de maior magnitude. Como o objetivo aqui era verificar a sensibilidade para mudanças maiores pré-especificadas, percebeu-se que os gráficos CUSUM individuais têm desempenho superior ao combinado.

No Capítulo 8 foi proposto um roteiro de passos para a construção de um gráfico combinado Shewhart-CUSUM binomial. Também foram avaliadas três aplicações do gráfico combinado. A primeira aplicação é para um conjunto de dados adaptado da literatura. As duas aplicações seguintes são em dados reais de dois processos, o monitoramento de um indicador de qualidade de água potável e o processo de rotulagem de garrafas em uma empresa de bebidas. Para complementar, em todas as aplicações os resultados dos gráficos combinados foram comparados com procedimentos CUSUM e Shewhart individuais. O gráfico combinado mostrou-se uma alternativa interessante para duas das três aplicações analisadas. Na situação proposta na primeira aplicação, sinalizou alterações grandes e temporárias na parte Shewhart e pequenas e persistentes na parte CUSUM. Na aplicação a dados reais no monitoramento de um indicador ambiental, o gráfico combinado não teve desempenho superior aos demais gráficos, em virtude das particularidades do processo. Mas para o processo de rotulagem de garrafas é uma opção, pois quando aumenta a sensibilidade do CUSUM para grandes alterações, permite sinalizar mudanças de várias magnitudes mantendo a taxa de alarmes falsos dos gráficos individuais.

Analisando os Capítulos 6 a 8 e, avaliando todos os gráficos estudados, não se pode dizer que um deles é melhor que o outro em todas as situações. Na maioria dos casos, nenhum gráfico especificamente terá uma performance superior em comparação a outros gráficos para todas as magnitudes de mudança (REYNOLDS e STOUMBOS, 2006). Assim um gráfico combinado para variáveis com distribuição binomial pode ser útil em processos específicos que necessitem:

- i. detectar rapidamente alterações pequenas a moderadas na proporção sob controle e
- ii. sinalizar também alterações maiores na proporção sob controle.

Assim, conclui-se que um gráfico combinado Shewhart-CUSUM binomial, com a adição das linhas Shewhart a um procedimento CUSUM binomial, aumenta a sensibilidade deste para alterações maiores que a especificada no seu planejamento (p_1). Se a finalidade é sinalizar especificamente alterações maiores, deve-se considerar um CUSUM planejado para tal, ou um gráfico do tipo Shewhart, conforme o caso.

9.1 CONTRIBUIÇÕES

Considerando-se os objetivos propostos e os resultados encontrados, este trabalho traz as seguintes contribuições científicas:

Foi desenvolvida uma nova abordagem para a aproximação do limite superior de um CUSUM binomial. É uma alternativa que busca tornar menos complexa e mais rápida esta tarefa. O modelo desenvolvido foi implementado computacionalmente no ambiente GNU R.

A partir de valores simulados do RL pode-se ter visualmente uma idéia da forma da sua distribuição e calcular alguns momentos, como valores de ARL. Estes valores, de ARL, podem ser utilizados para comparações em trabalhos futuros.

A avaliação do desempenho de um gráfico combinado comparando-o com gráficos CUSUM e do tipo Shewhart individuais permitiu confirmar que a adição de linhas Shewhart aumenta a sensibilidade de um gráfico CUSUM binomial para mudanças de maior magnitude. O que se tem é um gráfico sensível a uma faixa maior de magnitudes de mudança. Verificou-se que um gráfico combinado tem desempenho similar ao CUSUM nas proximidades da proporção p_1 de planejamento e superior na medida em que as alterações se tornam maiores, distanciando-se daquela de planejamento.

Identificou-se também que existe uma região onde o gráfico combinado sinaliza as alterações antes dos gráficos individuais. É uma

região de alterações magnitude moderada, maiores que as especificadas em projeto, e inferiores a aproximadamente 2σ em relação à média.

Foi elaborada de uma metodologia para construção de um gráfico combinado Shewhart-CUSUM binomial, envolvendo verificação das suposições das suposições necessárias, com sugestões de alternativas para o caso de estas não serem atendidas.

Rotinas no R para a construção de um gráfico CUSUM binomial e combinado Shewhart-CUSUM binomial foram elaboradas.

E, finalizando, uma contribuição importante foi a aplicação do gráfico combinado a dados adaptados da literatura e também a dados reais, o que não se tinha registros. Esta etapa foi relevante, pois permitiu transferir a situações práticas tudo o que foi desenvolvido neste trabalho.

9.2 DIFICULDADES ENCONTRADAS

Algumas deficiências foram sentidas no decorrer do trabalho, relacionadas a seguir:

- a) A distribuição binomial pode assumir características assimétricas, reduzindo o uso dos limites convencionais de Shewhart;
- b) Mesmo utilizando limites probabilísticos (ou exatos) nem sempre é possível alcançar o ARL_0 desejado;
- c) Como a distribuição é discreta os valores de ARL_0 , inclusive do CUSUM, estarão restritos a uma faixa de valores condicionada aos limites de controle;
- d) Em virtude das muitas especificidades de cada caso, tamanhos de amostra, valores de p_0 , p_1 e ARL_0 , não se conseguiu estabelecer faixas específicas tanto para a acurácia da aproximação como nas análises de desempenho realizadas.

9.3 SUGESTÕES PARA TRABALHOS FUTUROS

Como em todo trabalho, nem todas as possibilidades foram esgotadas. Assim, algumas sugestões para trabalhos futuros são apresentadas em seguida:

Na fórmula de aproximação do limite superior h :

a) investigar o comportamento do erro. Pode então ser viável incorporar este erro na Equação mediante, por exemplo, um argumento de ajuste de curvas;

b) aplicar um processo de interpolação mais sofisticado, como splines, que converge sempre, ao invés da interpolação logarítmica;

c) investigar os resultados sob a ótica de h e k como múltiplos de um mesmo número, como sugerido por Reynolds e Stoumbos (1998, 1999);

d) investigar outros processos de desenvolvimento em série da aproximação para h ;

e) desenvolver um algoritmo para a aproximação de h considerando ARL_0 e ARL_1 , avaliando em conjunto as Equações (3.32) e (3.37) (conforme Rogerson, 2006).

Com relação ao gráfico combinado:

a) efetuar novas análises, abrangendo mais valores para n ; p_0 ; p_1 e ARL_0 , procurando assim definir especificamente, a partir de que faixas de magnitude de mudança, o gráfico combinado tem desempenho superior ao CUSUM;

b) efetuar novas análises, abrangendo mais valores para n ; p_0 ; p_1 e ARL_0 , procurando assim definir de forma mais específica, as regiões onde o gráfico combinado tem desempenho superior aos dois gráficos, CUSUM e do tipo Shewhart; uma opção é termos de números de desvios padrão em relação à média (np_0) do processo;

c) comparar o desempenho do gráfico combinado a um CUSUM individual com RIR;

d) utilizar processos de simulação mais refinados, como por exemplo, técnicas de redução da variância e comparar os resultados;

e) investigar a aplicação de limites Shewhart com transformações, como a arco seno ou Box-Cox (ver Duran e Albin, 2009) ou outros limites modificados (Shore, 2000) (Wu e Luo, 2003) (Wu *et al.*, 2005);

f) abrangendo um gráfico bilateral;

g) avaliar o desempenho de um gráfico combinado com dados autocorrelacionados, aplicando limites alargados ou outros procedimentos que capturem a autocorrelação;

h) estudar o desempenho de um gráfico combinado em que os dados apresentem superdispersão, investigando opções para contornar este problema.

i) aplicar a metodologia desenvolvida no Capítulo 8 neste trabalho para a construção de gráfico combinado CUSUM-Shewhart binomial pode ser estendida a outras distribuições discretas, como por exemplo, a distribuição de Poisson.

Para finalizar, ressalta-se é fundamental conhecer o processo ao qual se deseja aplicar gráficos de controle. É preciso identificar o que deve ser monitorado e qual seria uma situação fora de controle estatístico que deve ser rapidamente detectada. Identificar que magnitudes de mudança interessam ser sinalizadas é essencial para a escolha adequada de um gráfico de controle. E, é claro, fazer com que as pessoas envolvidas com o processo participem e compreendam o trabalho desenvolvido.

REFERÊNCIAS

ABEL, V. On one-sided combined Shewhart-CUSUM control schemes for Poisson Counts. **Computational Statistics Quarterly**, v.6, p. 31-39, 1990.

ACOSTA-MEJIA, C. A. Improved p Charts to Monitor Process Quality. **IIE Transactions**, v. 31, p. 509 - 516, 1999.

ADLER D. **vioplot: Violin plot**. R package version 0.2.2005. Disponível em: <<http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler>>. Acesso em: 03 de abril de 2010.

ALVES, C. C. **Gráficos de Controle CUSUM: um enfoque dinâmico para a análise estatística de processos**. 2003. 119 f. Dissertação (Mestrado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2003.

ALVES, C. C. **Método da Equação integral com quadratura gaussiana para otimizar os parâmetros do gráfico de controle multivariado de somas acumuladas**. 2009. 196 f. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2009.

ALVES, C.C.; SAMOHYL, R.W. A utilização dos gráficos de controle CUSUM para o monitoramento de processos industriais. In: XXIV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO ENEGEP 2004, 2004, Florianópolis. **Anais - XXIV Encontro Nacional de Engenharia de Produção Enegep 2004**. Florianópolis, UFSC, 2004.

ALWAN, L. C.; ROBERTS, H.V. The Problem of Misplaced Control Limits. **Journal of Applied Statistics**, v. 44, p. 269-278, 1995.

BARRAT, Benjamim *et al.* Investigation into the use of the CUSUM technique in identifying changes in mean air pollution levels following introduction of a traffic management scheme. **Atmospheric Environment** 41, 1784-1791, 2007. Disponível em: <<http://www.elsevier.com/locate/atmosenv>>. Acesso em 05 nov. 2008.

BARTHELIS, Richard *et al.* A computational investigation of Conover's Kolmogorov-Smirnov test for discrete distributions. **Journal of Statistical Computation and Simulation**, v.7, n.2, p. 151- 161, 1978

BEILES, C. B.; MORTON, A. P. Cumulative sum control chart for assessing performance in arterial surgery. **ANZ Journal of Surgery**, v. 24, p. 146-151, 2004.

BENNET, B.M.; HSU, P. A sampling study of the power function of the binomial χ^2 “index of dispersion” test. *Journal of Hygiene*. v.59, n.4, p.449-455, 1961.

BERTHOUEX, P. M.; BROWN, L. C. **Statistics for Environmental Engineers**. 2. Ed. New York: Lewis Publishers, 2002. 463 p.

BEST, D.J.; RAYNER, J.C.W. Goodness of fit for the binomial distribution. **Australian Journal of Statistics**. v. 39, n. 3, p. 355-364, 1997.

BHAT, U.N.; LAL, R. Attribute Control Charts for Markov Dependent Production Processes. **IIE Transactions**,v.22, n.2, p.181-188, 1990.

BHAT, U. N.; LAL, R. Number of Success in Markov Trials. **Advanced Applied Probability**. v.20, p. 677-680, 1988.

BISSEL, A. F. Cusum Techniques for Quality Control. **Applied Statistics**. v. 18, n.1, p. 1-30, 1969.

BLACKSELL, Stuart *et al.* Use of combined Shewhart-CUSUM Control Charts in Internal Quality Control of Enzyme-linked immunosorbent assays for the typing of foot and mouth disease virus antigen. **Revue Scientifique et Technique (International Office of Epizootics)**, v.13, n.3, p. 687-699, 1994.

BLACKSELL, Stuart *et al.* Implementation of internal laboratory quality control procedures for the monitoring of ELISA performance at a regional veterinary laboratory. **Veterinary Microbiology**, v.51, p. 1-9, 1996.

BOURKE, P. D. Detecting a shift in Fraction Nonconforming Using Run-Length Control Charts with 100% Inspection. **Journal of Quality Technology**, v. 23, n. 3, p. 225-238, 1991.

BOURKE, P. D. Sample size and the Binomial CUSUM Control Chart: the case of 100% inspection. **Metrika**, v.53, p. 5-70, 2001.

BOX, G. P., COLEMAN, D.E., BAXLEY, R.V. A comparison of statistical process control and engineering process control. **Journal of Quality Technology**.v.29, p. 128 -130, 1997.

BRASIL. **Ministério da Saúde**. Secretaria de Vigilância em Saúde, Coordenação-Geral de Vigilância em Saúde Ambiental. Portaria no. 518/2004 do Ministério da Saúde.Disponível em:<http://portal.saude.gov.br/portal/arquivos/pdf/portaria_518_2004.pdf>
Acesso em 20/08/2010.

BRENT, R. **Algorithms for Minimization without Derivatives**. Englewood Cliffs, NJ: Prentice-Hall, 1973. Disponível em: <<http://www.netlib.org/c/brent.shar>>. Acesso em 20/08/2010.

BROADBENT, S. R. The Inspection of a Markov Process. **Journal of Royal Statistical Society Series B**.v. 20, p.111-119, 1958.

BROOK, D.; EVANS, D. A. An Approach to the Probability Distribution of Cusum Run Length. **Biometrika**, v. 59, n. 3, p. 539-549, 1972.

CAMERON, A. C.; TRIVEDI, P.K. Essentials of Count Data Regression. In: BALTAGI, B.H. **A Companion to Theoretical Econometrics**. Blackwell Publishing, 2003. p. 331-348.

CHANG, J. T. On Moments of the First Ladder Height of Random Walks with Small Drift. **The Annals of Applied Probability**, v. 2, n. 3, p. 714-738, 1992.

CHANG, T.C.; GAN, F.F. Cumulative Sum Charts for high yields processes. **Statistica Sinica**,v.11, p.791-805, 2001.

CHAPEAU-BLONDEAU, F.; MONIR, A. Numerical evaluation of the Lambert W function and applications to generation of generalized Gaussian noise with exponent $\frac{1}{2}$. **IEEE Transactions Signal Processing**, v. 50, p. 2160 – 2165, 2002.

CHOU, C. J.; O'BRIEN, R.F.; BARNETT. D. B. Application of intrawell testing of RCRA groundwater monitoring data when no upgradient well exists. **Environmental Monitoring and Assessment**, n. 71, p. 91-106, 2001.

COELHO, L.C. **Utilização de modelos de suavização exponencial para previsão de demanda com gráficos de controle combinados Shewhart-CUSUM**. 2003. 121f. Dissertação de Mestrado (Mestrado em Engenharia de Produção) -Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2008.

CONRATHS, F.J.; SCHARES, G. Validation of Molecular-diagnostic techniques in the parasitological laboratory. **Veterinary Parasitology**, v. 136, n.2, p.91-98, 2006.

CORLESS, Robert *et al.* On the LambertW function. **Advances in Computacional Mathematics**, v. 5, p.329-359, 1996.

COSTA, A. F. B. EPPRECHT, E. K. CARPINETTI, L.C.R. **Controle Estatístico de Qualidade**. São Paulo: Editora Atlas S.A, 2004. 334 p.

COULQUHOUN, P. H. D. CUSUM analysis of J-pouch surgery reflects no learning curve after board certification. **Canadian Journal of Surgery**, v. 51, n.4, p. 296-299, 2008.

CUI, H.; LUND, R. A new look at time series of counts. **Biometrika**. v. 96, n. 4, p. 781-792, 2009.

DEVORE, J. L. **Probabilidade e Estatística para Engenharia e Ciências**. 6 ed. São Paulo: Pioneira Thomson Learning, 2006. 692 p.

DU, J.; LI, Y. The integer-valued autoregressive (INAR(p)) model. **Journal of Time Series Analysis**, v.12, p. 129-42, 1991.

DUNCAN, A.J. **Quality Control and Industrial Statistics**. 5 ed. Irwin Homewood IL, 1986. 1296 p.

DURAN, R. I.; ALBIN, S. L. Monitoring a Fraction with Easy and Reliable Settings of the False Alarm Rate. **Quality and Reliability Engineering International**, v. 25, p. 1029-1043, 2009.

EPPRECHT, E. K.; COSTA, A. F.B. Adaptive Sample Size Control Charts for Attributes. **Quality Engineering**, v.13, n.3, p. 465-473, 2001.

EWAN, W.D.; KEMP, K.W. Sampling Inspection of Continuous Processes with no Autocorrelation Between Successive Results. **Biometrika**, v. 47, n. 3/4, p. 363-380, 1960.

FRISEN, M. Properties and Use of the Shewhart Method and Followers. **Sequential Analysis**, v.26, p. 171-193, 2007.

GAN, F. F. An optimal design of CUSUM quality control charts. **Journal of Quality Technology**, v.23, p.279-286, 1991.

GAN, F. F. An optimal design of CUSUM control charts for binomial counts. **Journal of Applied Statistics**, v. 20, n. 4, p. 445-460, 1993.

GAUTHIER, G.; LATOUR, A. Convergence forte des estimateurs des paramètres d'un processus GENAR (p). **Annales des sciences mathématiques du Québec**. v. 18, p. 37-59, 1994.

GIBBONS, R. D. Use of Combined Shewhart-CUSUM Control Charts for Ground Water Monitoring Applications. **Ground Water**, v.5, n. 37, p. 682 - 692, set./out. 1999.

HAWKINS, D. M. Evaluation on Average Run Lengths of Cumulative Sum Charts for an Arbitrary Data Distribution. **Communication in Statistics – Simulation and Computation**, v. 21 (4), p.1001-1020, 1992.

HAWKINS, D. OLWELL, D. **Cumulative Sum Charts and Charting for Quality Improvement**. New York: Springer Verlag Inc., 1998. 247 p.

HENNING, Elisa *et al.* Análise de um projeto de ensino utilizando Gráfico de Controle de Soma Cumulativa. In: 37 Congresso Brasileiro de Educação em Engenharia - COBENGE 2009, 37., 2009, Recife. **Anais...Cobenge 2009**. Recife: Abenge, 2009.

HINTZE, J.L.; RAY D. N. Violin Plots: A Box Plot-Density Trace Synergism. **The American Statistician**. v.52, n.2, p.181-184, 1998.

HÖHLE, M. Surveillance: An R package for the monitoring of infectious diseases. **Computational Statistics**, v. 22, p. 571 – 572, 2007.

JIAO, J.X.; HELO, P.T. Optimization design of CUSUM control chart based on Taguchi's loss function. **International Journal of Advanced Manufacturing Technology**, v. 35, p. 1234 – 1243, 2006.

JOHNSON, N.L.; KEMP, A.W.; KOTZ, S. **Univariate Discrete Distributions**. 3 ed. New Jersey:John Wiley and Sons , 2005. 646 p.

JOHNSON, N.L.; LEONE, S.C. Cumulative Sum Control Charts, Parts I and III. **Industrial Quality Control**, v.17, p.15-21; e v.19; pp. 22-28, 1962.

JONES, G.; GOVINDARAJU, K. A Graphical Method for Checking Attribute Control Charts Assumptions. **Quality Engineering**. v.13, n.1, p.19-26, 2000.

JUN, C.; CHOI, M. S.; Simulating the Average Run Length for CUSUM Schemes using Variance Reduction Techniques. **Communication in Statistics - Simulation and Computation**, v. 22, n.3, p.877-887, 1993.

JUNG, R.; TREMAYNE, A. R. Coherent Forecasting in integer time series models. **International Journal of Forecasting**. v. 22, p.223-238, 2006.

JURAN, J. M.; GRYNA, F.M. **Controle da Qualidade**. 4a. ed. São Paulo: Makroon Books, 1992.

KEMP, K. W. The Use of Cumulative Sums for Sampling Inspection Schemes. **Applied Statistics**, v. 11, n. 1, p. 16-31, 1962.

KHOO, M. B. C. A moving average control chart for monitoring the fraction non-conforming. **Quality Reliability Engineering International**, v.20, p. 617 – 635, 2004.

KIM, S. H.; ALEXOPOULOS, C.; TSUI, K. L; WILSON, J. R. A distribution-free tabular CUSUM chart for autocorrelated data. **IIE Transactions**, v. 39, p. 317 – 330, 2007.

KNOTH, S. **SPC: Statistical Process Control**. R package version 0.2.1 (2009). Disponível em: <<http://CRAN.R-project.org/package=spc>>. Acesso em 02/12/2009.

LAI, C. D.; GOVINDARAJU, K.; XIE, M. Effects of Correlation on Fraction Nonconforming Statistical Process Control Procedures. **Journal of Applied Statistics**. v. 25, n. 4, p.535-543, 1998.

LAI, C. D., GOVINDARAJU, K.; XIE, M. Study of a Markov Model for a High-Quality Dependent Process. **Journal of Applied Statistics**. v.27, n.4, p.461-473, 2000.

LATOUR, A. Existence and Stochastic Structure of a Non-negative Integer-Valued Autoregressive Process. **Journal of Time Series Analysis**, v. 19, n. 4, p. 439-455, 1998.

LUCAS, J. M. Combined Shewhart-CUSUM Quality Control Schemes. **Journal of Quality Technology**, v. 14, n. 2, p. 51-59, 1982.

LUCAS, J. M.; CROISIER, R. B. Fast Initial Response for CUSUM Quality Control Schemes. **Technometrics**, v. 24, n. 3, p. 199-205, 1982.

LUCAS, J. M. Counted Data CUSUM's. **Technometrics**, v. 27, n. 2, p. 129-144, 1985.

LUCAS, J. Control schemes for low count level. **Journal of Quality Technology**, v.21, p. 109 – 201, 1989.

LUCAS, J. M.; DAVIS, D. J.; SANIGA, E. M. Detecting improvement using Shewhart attribute control charts when the lower control limit is zero. **IIE Transactions**, v.38, p. 659-669, 2006.

MADSEN, R. W. Generalized Binomial Distributions. **Communications in Statistics: Theory and Methods**. Vol. 22, p. 3065-3086, 1993.

MASON, R. L.; YOUNG, J. C. Multivariate Statistical Process Control with Industrial Applications. **ASA/SIAM**: Philadelphia, PA, 2002.

MESNIL, B.; PETITGAS, P. Detection of changes in time-series of indicators using CUSUM control charts. **Aquatic Living Resources**, v. 22, p. 187–192, 2009.

MEYER, D.; ZEILEIS, A.; HORNIK, K. **vcd: Visualizing Categorical Data**. R package version 1.2-7, 2009. Disponível em: <<http://www.r-project.org>>. Acesso em 06/07/2010.

MEYER, N. *et al.* A multi-data source surveillance system to detect a bioterrorism attack during the G8 Summit in Scotland. **Epidemiology and Infection**, v. 136, p. 876–885, 2008.

MINGOTI, S. A.; CARVALHO, J. P. A cadeia de Markov no monitoramento, via inspeção por atributos, de processos autocorrelacionados: o programa SCAP- módulo AI. In: XXIII ENEGEP Encontro Nacional de Engenharia de Produção, 2007, Ouro Preto. **Anais.Ouro Preto,XXIIIENEGEP**, 2003.

MONTGOMERY, D. C. **Introdução ao Controle Estatístico de Qualidade**. 4. ed. Rio de Janeiro: LTC, 2004. 513 p.

MORAIS, M. C. **Ordenação estocástica na análise de desempenho de esquemas de controle de qualidade**. 2002. 194f. Tese (Doutorado em Matemática) - Instituto Superior Técnico, Universidade Técnica de Lisboa, 2002. Disponível em:<<http://www.math.ist.utl.pt/~mjmorais/thesis.pdf>> . Acesso em 19/09/08.

MORAIS, M. C. **Fiabilidade e Controle de Qualidade: Material Didático –Esquemas de controlo de qualidade do tipo CUSUM e EWMA para atributos e variáveis**.2004. 26 f.Notas de aula. Disponível em: <http://www.math.ist.utl.pt/mmorais/FCQ0-0607.cap10.pdf>. Acessado em: 01/09/2007.

MORAIS, M. C.; PACHECO, A. Combined CUSUM-Shewhart Schemes for Binomial Data. **Economic Quality Control**. v. 21, n.1, p. 43-57, 2006.

MORTON, Anthony et al. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. **Journal of Quality in Clinical Practice**, v. 21, p. 112-117, 2001.

MOSTELLER, F.; TUKEY, J. W. The Uses and Usefulness of Binomial Probability Paper. *Journal of the American Statistical Association*. Vol 44, p. 174 - 212, 1949. In: **Selected Papers of Frederik Mosteller**. Springer Series in Statistics, 2006. p. 111-145.

MOUSTAKIDES, G. V. Optimal Stopping Times for Detecting Changes in Distributions. **The Annals of Statistics**, v. 14, n. 4, p. 1379-1387, dez. 1986.

MUNFORD, A. G. A Control Chart based on Cumulative Scores. **Applied Statistics**, v. 29, n.3, p. 252-258, 1980.

NAIK, V. N.; DE VITO, I.; HALPERN, S. H. Cusum Analysis is a useful tool to asses resident proficiency at insertion of labour epidurals. **Canadian Journal of Anesthesiology**, v. 50, n.7, p. 694-698, 2003.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. U.S. Departments of Commerce. **Dataplot reference manual: Violin plot**.2003. Disponível em:
<<http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/violplot.htm>> Acesso em 22/08/2010.

NELSON, L. S. A. Supplementary Runs Tests for np Control Charts. **Journal of Quality Technology**, v. 29, n.2, p. 225-227, 1994.

ÖZDEMİR, T.; EYDURAN, E. Comparison of Chi-Square and Likelihood Ratio Chi-Square Tests: Power of Test. **Journal of Applied Sciences Research**, v.1, n.2, p. 242-244, 2005.

PAGE, E. S. Continuous Inspection Schemes. **Biometrika**, v. 41, n. 1/2, p. 100-114, Jun. 1954.

PALADINI, E. P. **Gestão da Qualidade**. São Paulo: Atlas, 2000. 330 p.

RADAELLI, G. Poisson and negative binomial dynamics for counted data under CUSUM-type charts. **Journal of Applied Statistics**, v. 21, p. 347 – 356, 1994.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2009. Disponível em: <<http://www.R-project.org>>.

REYNOLDS, M.R; STOUMBOS, Z.G. The SPRT chart for monitoring a proportion. **IIE Transactions**, v. 30, p. 545-561, 1998.

REYNOLDS, M. R.; STOUMBOS, Z. G. A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously. **Journal of Quality Technology**, v. 31, no.1, Pag 87-108, Jan 1999.

REYNOLDS, M. R.; STOUMBOS, Z. G. A general approach to modeling CUSUM charts for a proportion. **IIE Transactions**. v. 32, p. 515-535, 2000.

REYNOLDS, M. R.; STOUMBOS, Z. G. Monitoring a Proportion using CUSUM and SPRT charts. In: LENZ, H. J.; WILRICH, P.T. **Frontiers in Statistical Quality Control**. Physika-Verlag Heidelberg: 2001. p. 155-175.

REYNOLDS JR., M. R., STOUMBOS, Z. G. Comparisons of some exponentially weighted moving average control charts for monitoring the process mean and variance. **Technometrics**, v.48 , 550-567, 2006.
RIZZON, L. A.; MENEGUZZO, J. Sistema de Produção de Vinho Tinto. **Embrapa Uva e Vinho. Sistemas de Produção**, v.12, 2006. Disponível em

<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Vinho/SistemaProducaoVinhoTinto/engarrafamento.htm>. Acesso em 05/07/2010.

ROCHA, R. **Implementação de sistema gerencial, com avanços em controle estatístico, em laboratório de nutrição animal**. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2004.

ROGERSON, P. A. Formulas for the Design of CUSUM Quality Control Charts. **Communications in Statistics – Theory and Methods**, n.35, 373-383, 2006.

ROSA, J. M. C. **Modelos INAR(1) e Estruturais para Séries Temporais de Contagens: Um estudo comparativo, utilizando as distribuições Poisson e Geométrica**. Dissertação de Mestrado (Mestrado em Estatística) – Instituto de Matemática e Estatística Universidade Estadual de Campinas, Campinas, 1997.

ROSS, S. M. Variance reduction in simulation via random hazards. **Probability in the Engineering and Informational Sciences**, v.4, p. 299-309, 1990.

RUSSO, S. L. **Gráficos de controle para variáveis não conformes autocorrelacionadas**. 2002. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2002.

SAMOHYL, R.W. Controle Estatístico de Qualidade. Rio de Janeiro: Elsevier, 2009. 275 p.

SANCHES O. Princípios básicos de procedimentos estatísticos aplicados na análise de dados de vigilância em saúde pública: uma

revisão. **Cadernos de Saúde Pública**, Rio de Janeiro, v.16, n.2, p.317-333, abr/jun 2000.

SANIGA, E. M.; DAVIS, D.J.; MCWILLIANS, T.P. Economic, Statistical, and economic statistical design of attribute charts. **Journal of Quality Technology**, v.27, p. 56-73, 1995.

SAVIC, M. P. Charts in the Quality Control of the Grading Process in the High Education. **Panoeconomicus**, n. 3, p. 335-347, 2006.
Disponível em: <<http://www.panoeconomicus.rs/casopis/trecibroj>>
Acesso: em 02 jan. 2009.

SCHADER, M.; SCHMID, F. Two rules of thumb for the approximation of the binomial distribution by the normal distribution. **The American Statistician**, v.43, n.1, p. 34- 24, 1989.

SCHNEIDER, H.; O' CINNEIDE, C. Design of CUSUM Control Charts Using Narrow Limits Gauges. **Journal of Quality Technology**. v. 19, n. 2, 1987.

SCRUCCA, L. qcc: An R package for quality control charting and statistical process control. **R News**, v. 4, n.1, p. 11-17, 2004. Disponível em: <<http://www.r-project.org>. >. Acessado em 11/08/2005.

SHARMA, P. R. **A study of properties and applications of control charts for high yield processes**. Tese (Doctor Of Philosophy) - Department of Industrial e Systems Engineering, The National University Of Singapore, 2003.

SHORE, H. General control charts for attributes. **IIE Transactions**, v.32, 1149-1160, 2000.

SIEGMUND, D. Corrected diffusion approximations in certain random walks problems. **Advances in Applied Probability**, v.11, p. 701 – 719, 1979.

SIEGMUND, D. **Sequential Analysis: Test and Confidence Intervals**, Springer Verlag, New-York, 1985.272 p.

SILVA, I.; SILVA, M. E. Seleccção de ordem em modelos INAR. **Actas do X Congresso Anual da Sociedade Portuguesa de Estatística**. Edições SPE: Lisboa, p. 1-11, 2002.

SILVA, I. M. M. **Contributions to the analysis of discrete-valued time series**. Tese (Doutorado em Matemática) - Departamento de Matemática Aplicada, Faculdade de Ciências da Universidade do Porto, 2005. Disponível em:
<http://paginas.fe.up.pt/~ims/phd_thesis_Isabel_Silva.pdf.>Acesso em 20/08/2010.

SOLBERG, H. Monitoring long-term analytical quality by computerized combined Shewart-cusum method. **Scandinavian Journal of Clinical and Laboratory Investigation**. v. 172, p. 43-49, 1984.

SONESSON, C.; BOCK, D. A review and discussion of prospective statistical surveillance in public health. **Journal of the Royal Statistical SocietyA**. v. 166, no. 1, p. 5-21, 2003.

SOUZA, G. P.; SAMOBYL, R. W. **Monitoring Forecast Errors with Combined CUSUM and Shewhart Control Charts**. Proceedings of the 26th. International Symposium of Forecasting, 2008.

STOUMBOS, Zachary G. *et al.* The State of Statistical Process Control as We Proceed into the 21st Century. **Journal of the American Statistical Association**, v.95, n. 451, p. 992-998, 2000.

TRINDADE, A. L. G. **Contribuições para o controle on-line de processos por atributos**. 2008. Tese (Doutorado em Engenharia de Produção) – Escola Politécnica da Universidade de São Paulo, Universidade de São Paulo, São Paulo, 2008.

TSUNG, F.; WANG, K. Adaptive Charting Techniques: Literature Review and Extensions. In: LENZ, H.J. WILRICH, P. SCMD, W. **Frontiers in Statistical Quality Control**, Vol 9. Springer Physika-Verlag, 2010. p. 19-36.

USEPA. **Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance**. Washington: Office of Solid Waste, 1989. Disponível em <http://www.epa.gov/otaq/transp/statanal.pdf>. Acesso em 10 de setembro de 2008.

VIEIRA, S. **Estatística para a Qualidade**. Rio de Janeiro: Elsevier, 1999. 198 p.

XIE, M.; GOH, T. N., SPC of A Near Zero-defect Process Subject to Random Shock. **Quality and Reliability Engineering International**, v.9, p. 89-93, 1993.

WALD, A. Sequential Tests of Statistical Hypothesis. **Annals of Mathematical Statistics**. Vol 16, n.2, p. 117-186, jun. 1945.

WALD, A. **Sequential Analysis**. Dover Publications, New York, 1947. 224p.

WALD, A.; WOLFOWITZ, J. An Exact Test for Randomness in the Non-parametric Case Based on Serial Correlation. **Annals of Mathematics and Statistics**, v. 14, n. 4, p. 378-388, 1943.

WEIß, C. H. Controlling correlated processes of Poisson counts. **Quality and Reliability Engineering International**, v. 23, n. 6, p.741-754, 2007.

WEIß, C. H.; TESTIK, M. C. CUSUM Monitoring of First-Order Integer-valued Autoregressive Processes of Poisson Counts. **Journal of Quality Technology**, v. 41, n. 4, p. 389 - 400, 2009.

WEIß, C. H. EWMA Monitoring of Correlated Processes of Poisson Counts. **Quality Technology e Quantitative Management**, v. 6, n. 2, p. 137-153, 2009

WESTGARD, James *et al.* Combined Shewhart-CUSUM Control Chart for Improved Quality Control in Clinical Chemistry. **Clinical Chemistry**, v. 23, n. 10, p. 1881-1887,1977.

WETHERILL, G. B.; BROWN, D. W. **Statistical Process Control: Theory and Practice**. Chapman and Hall, New York, 1991. 400 p.

WOODALL, W. H.; ADAMS, B. M. The Statistical Design of CUSUM Charts. **Quality Engineering**, v. 5, p. 559-570, 1993.

WOODALL, W. H.; NCUBE, M. M. A Combined Shewhart-Cumulative Score Quality Control Chart. **Applied Statistics**, v. 33, n. 3, p. 259-265, 1984.

WOODALL, W.H. Control Charts based in Attribute Data: Bibliography and Rewiew. **Journal of Quality Technology**, v.29, p.172-183, 1997.

WOODALL, W. H. The Use of Control Charts in Health-Care and Public-Health Surveillance. **Journal of Quality Technology**, v. 38, n. 2, p. 89-104, 2006.

WU, Z.; LUO, H. Three-triplet np control charts. **European Journal of Operational Research**, v.149, p.614-624, 2003.

WU, Z.; YEO, S. H.; SPEDING, T. A. A synthetic control chart for detecting fraction nonconforming increases. **Journal of Quality Technology**, v.33, p.104 -111, 2001.

WU, Z.; LUO, H.; ZHANG, X. L. Optimal np chart with curtailment. **European Journal of Operational Research**, v.174, pp.1723-1741, 2006.

WU, Zhang *et al.* Optimization designs of the combined Shewhart-CUSUM control charts. **Computational Statistics and Data Analysis**, v. 53, p. 496 - 506, 2008.

WU, Z.; JIAO, J.; LIU, Y. A binomial CUSUM chart for detecting large shifts in fraction non conforming. **Journal of Applied Statistics**. v. 35, n.11, p. 1267-1276, 2008.

WU, Y. **Inference for change-point and post-change means after a CUSUM test**. Lecture Notes in Statistics. New York: Springer Verlag, 2005. 158 p.

YASHCHIN, E. On the Analysis and Design of CUSUM-Shewhart Control Schemes. **IBM Journal of Research and Development**. v. 29, n.4, p. 377-391, 1985.

ZHOU, Wanfang *et al.* Application of water quality control charts to spring monitoring in karst terranes. **Environmental Geology**, v. 53, no. 6, 2008.

APÊNDICE A – Rotina no R para aproximação do limite superior de um CUSUM binomial usando a função LambertW

```

###funcao para aproximar h com LambertW na resolucao
para b (h+correcao)###

hrs.l<-function(p0,p1,n,arl0) {
r1<- -log((1-p1)/(1-p0))
r2<- log((p1*(1-p0))/(p0*(1-p1)))
g <- r1/r2
k <- round(n*g,2)
l1<-log((exp(-((n *(r2*g + (-1 + exp(n* r2*g)) *(r1 -
p0* r2)* arl0)))/(-1 + exp(n* r2 *g)))* n* r2 *g)/(-1
+ exp(n* r2*g))
l2<-log(-log((exp(-((n *r2* g)/(-1 + exp(n* r2* g)))
+ n*(-r1 + p0* r2)* arl0)* n *r2 *g)/(-1 + exp(n* r2
*g))))
l3<-l2/l1
l4<- (l2*(l2-2))/(2*l1^2)
l5<- (l2*(6-9*l2+2*l2^2))/(6*l1^3)
b<-((-n*r2*g)/(1-exp(-n*r2*g))-(l1-l2)+n*arl0*(p0*r2-
r1))/r2 # 2 termos
b1 <- ((-n*r2*g)/(1-exp(-n*r2*g))-(l1-
l2+l3)+n*arl0*(p0*r2-r1))/r2 # 3 termos
b2 <- ((-n*r2*g)/(1-exp(-n*r2*g))-(l1-
l2+l3+l4)+n*arl0*(p0*r2-r1))/r2 # 4 termos
b3 <- ((-n*r2*g)/(1-exp(-n*r2*g))-(l1-
l2+l3+l4+l5)+n*arl0*(p0*r2-r1))/r2 # 5 termos
h<- b-(1/3)*(1-2*p0)
h1<- b1-(1/3)*(1-2*p0)
h2<- b2-(1/3)*(1-2*p0)
h3<- b3-(1/3)*(1-2*p0)
structure(list(k=k,h=h,h1=h1,h2=h2,h3=h3))
}

```

**APÊNDICE B – Rotina no R para aproximação do limite superior
de um CUSUM binomial com a aproximação CD, resolvendo a
Equação com um método iterativo**

```
###função para aproximar h resolução da Equação por
um método numérico

hrs<-function(p0,p1,n,arl0){
  h.rs<-uniroot(function (h) arl0-((log((p1*(1-
p0))/(p0*(1-p1))))/(n*((log((p1*(1-p0))/(p0*(1-
p1))))*p0-(-log((1-p1)/(1-p0)))
)))*((h+(1/3)*(1-2*p0))-n*((-log((1-p1)/(1-
p0))))
/(log((p1*(1-p0))/(p0*(1-
p1)))))*exp((h+(1/3)*(1-2*p0))*log((p1*(1-
p0))/(p0*(1-p1))))-1)/(1-exp(-n*((-log((1-p1)/(1-
p0))))
/(log((p1*(1-p0))/(p0*(1-p1)))))*log((p1*(1-
p0))/(p0*(1-p1))))),c(0,120))
  return(h.rs)
}
```

APÊNDICE C – Roteiro e Rotina no R para as simulações do Capítulo 5

ENTRADA:

- a) N = número de amostras (*valores de N*);
- b) n = tamanho da amostra;
- c) p_0 = probabilidade considerada sob controle;
- d) p_1 = probabilidade considerada fora de controle estatístico que se deseja detectar rapidamente;
- e) p = probabilidade ($p_0, sp_0 \dots$) com $s(\%)$;
- f) k = valor de referência do CUSUM;
- g) h = limite superior do CUSUM (combinado);
- h) r = número de replicações;

PASSOS:

- a) Simular 100 amostras aleatórias de tamanho n com distribuição binomial e probabilidade $p = p_0$ (alarme falso);
- b) Simular N amostras aleatórias de tamanho n com distribuição binomial e probabilidade $p = p_1$ (alarme verdadeiro);
- c) Aplica rotina do gráfico combinado - fl.2;
- d) Registra o número da primeira amostra onde ultrapassa o limite superior h ;
- e) Repete o procedimento r vezes;

Calcula a média (ARL), mediana (MRL), demais quartis e desvio padrão dos r valores para cada gráfico.

SAÍDA

- a. ARL, MRL, q_1 , q_3 , desvio padrão (SDRL);
- b. Planilha com os valores.

OBS: Para as simulações cujos valores estão na Tabela 5.6 considerou-se 30 amostras sob controle ($p = p_0$)

```

####rotinas simulacao - capta o ponto em que
ultrapassa h

####100 OBSERVACOES SOB CONTROLE####
m1<-rbinom(100,prob=p0,size=n)
mean(m1)

f1.2<-function (y,k,h){
zn=0
for (n in 1:length(y)){zn[n]=max(0,zn[n-1]+y[n]-k)}
x<-1:length(y)
pontos<-matrix(c(0,0),ncol=2,nrow=length(y))
for (n in 1:length(y)){pontos[n,1]=x[n]}
for (n in 1:length(y)){pontos[n,2]=zn[n]}
pontos.fora<-subset(pontos,zn>h)
pontos.dentro<-subset(pontos,zn <=h)
ponto<-
ifelse(nrow(pontos.dentro)>=nrow(pontos),NA,pontos.fora[1,1])
structure(list(ponto=ponto))}

ponto.det<-matrix(c(0),ncol=1,nrow=2000)
m3<-matrix(c(0),ncol=2000,nrow=20100)
for (i in 1:2000) {
m2<-rbinom(20000,size=50,prob=p1[1])
m3[,i]<-c(m1,m2)}
for (i in 1:2000){
ponto.det[i,1]<-f1.2(m3[,i],k=k.val,h=h.val)$ponto}

ponto.det.s<-ponto.det-100
hist(ponto.det.s)
summary(ponto.det.s)
round(median(ponto.det.s,na.rm=TRUE),0)
mad(ponto.det.s,na.rm=TRUE)
sd(ponto.det.s,na.rm=TRUE)

```


APÊNDICE D – Rotina no R para aproximação do limite superior de um CUSUM binomial, com a função LambertW e interpolação logarítmica

```

###função para calcular h com LambertW, cadeias de
Markov e interpolação logarítmica

hrs.5<-
function(p0,p1,n,ar10,tol=0.01,digits=1,digits.h=1)
{library(surveillance)
r1<- -log((1-p1)/(1-p0))
r2<- log((p1*(1-p0))/(p0*(1-p1)))
g <- r1/r2
k <- round(n*g,digits)
l1<-log((exp(-((n *(r2*g + (-1 + exp(n* r2*g)) *(r1 -
p0* r2)* ar10)))/(-1 + exp(n* r2 *g)))* n* r2 *g)/(-1
+ exp(n* r2*g))
l2<-log(-log((exp(-((n *r2* g)/(-1 + exp(n* r2* g))
+ n *(-r1 + p0* r2)* ar10)* n *r2 *g)/(-1 + exp(n* r2
*g))))
l3<-l2/l1
l4<-(l2*(l2-2))/(2*l1^2)
b2 <- ((-n*r2*g)/(1-exp(-n*r2*g)))-(l1-
l2+l3+l4)+n*ar10*(p0*r2-r1)/r2 # 4 termos
h<- round(b2-(1/3)*(1-2*p0),digits.h)
ar1<-
ar1Cusum(k=k,h=h,n=n,theta=p0,digits=digits,distr=c("
binomial"))$ARL
erp.1<-abs(ar1-ar10)/ar10
if (erp.1 < tol) {return(c(k=k,h=h,ARL=ar1))}
else {erp.2<-erp.1
hi <- h
while (erp.2 > tol){h1<- hi+1
h2<- hi-1
ar11<-
ar1Cusum(k=k,h=h1,n=n,theta=p0,digits=digits,distr=c(
"binomial"))$ARL
ar12<-
ar1Cusum(k=k,h=h2,n=n,theta=p0,digits=digits,distr=c(
"binomial"))$ARL
val<-function(f, x0, x1, f0, f1) {
x0 + ((x1 - x0) * (log(f) - log(f0)))/(log(f1) -
log(f0))}
h<-round(val(ar10,h1,h2,ar11,ar12),1)

```

```
arli<-  
arlCusum(k=k,h=h,n=n,theta=p0,digits=digits,distr=c("  
binomial"))$ARL  
erp.2<-abs(arli-arl0)/arl0  
hi <- h}  
return(c(k=k,h=hi,ARL=arli))}}
```

APÊNDICE E – Roteiro e Rotina no R para as simulações do Capítulo 6

ENTRADA:

- a) N = número de amostras (*valores de N*);
 $N = (500, 1000, 2000, 5000, 10000, 20000)$;
- b) n = tamanho da amostra;
- c) p_0 = probabilidade considerada sob controle;
- d) p_1 = probabilidade considerada fora de controle estatístico que se deseja detectar rapidamente;
- e) p = probabilidade ($p_0, sp_0 \dots$) com $s(\%)$;
- f) k = valor de referência do CUSUM;
- g) $h.val$ = limite superior do CUSUM (combinado);
- h) $h.c$ = limite superior do CUSUM;
- i) LSC = limite superior de Shewhart ou probabilístico (combinado);
- j) $LSCS$ = limite superior de Shewhart ou probabilístico (individual);
- k) r = número de replicações.

SIMULAÇÕES

1. Simular N amostras aleatórias de tamanho n com distribuição binomial e probabilidade $p = p_1$ (alarme verdadeiro);
2. Aplica rotina do gráfico combinado - COMBBINOM;
3. Registra o número da primeira amostra onde ultrapassa o limite superior h ;
4. Registra o número da primeira amostra onde ultrapassa o LSC ;
5. Seleciona e guarda o menor dos dois valores anteriores
 - a. Este é o ponto (valor) a ser analisado;
6. Aplica a rotina do gráfico CUSUM binomial - $f1$;
7. Registra o número da primeira amostra onde ultrapassa o limite superior h ;
8. Aplica a rotina do gráfico do tipo SHEWHART – $f2$;
9. Registra o número da primeira amostra onde ultrapassa o limite superior $LSCS$;
10. Repete o procedimento r vezes;

Calcula a média (ARL), mediana (MRL), demais quartis, desvio padrão, coeficiente de assimetria e curtose dos r valores para cada rodada
Plota gráficos – histogramas, violin plot e boxplot

SAÍDA

- a) ARL, MRL, $q1$, $q3$, desvio padrão (SDRL);
- b) Gráficos;
- c) Planilha com os valores;

São analisadas as situações

- a) Emissão de alarme falso ($p = p_0$);
- b) Rapidez na detecção da probabilidade $p_1 (p = p_1)$;
- c) Rapidez na detecção das probabilidades $p > p_0$ e $p \neq p_1$, com pequenas, moderadas e grandes alterações.

```
##SIMULACOES replica r vezes as rotinas acima, faz
histograma e calcula medidas
N=valor
r=20000
#combinado
ponto.det<-matrix(c(0),ncol=3,nrow=r)
for (i in 1:r) {
y<-rbinom(N,size=n,prob=p1.val)
ponto.sc<-combbinom3.3(y,n=n,k=k.val,h=h.val,lsc=lsc)
ponto.det[i,1]<-ponto.sc$ponto.cusum
ponto.det[i,2]<-ponto.sc$ponto.shewhart
ponto.det[i,3]<-ponto.sc$ponto.fora}
ponto.det[1:10,]
par(mfrow=c(1,3))
hist(ponto.det[,3],main=paste("COMB.
p=",p1.val),col="gray")
hist(ponto.det[,1],main=paste("CUSUM
p=",p1.val),col="white")
hist(ponto.det[,2],main=paste("SHEWHART
p=",p1.val),col="white")
par(mfrow=c(1,1))
kurtosis(ponto.det[,3],method=c("moment"))
skewness(ponto.det[,3],method=c("moment"))
sd(ponto.det[,3],na.rm=TRUE)
summary(ponto.det[,3])
summary(ponto.det[,1])
summary(ponto.det[,2])
quantile(ponto.det[,3],probs=c(5,10,25,50,75,95)/100)
quantile(ponto.det[,1],probs=c(5,10,25,50,75,95)/100)
quantile(ponto.det[,2],probs=c(5,10,25,50,75,95)/100)
vioplot(ponto.det[,1],ponto.det[,2],ponto.det[,3],col
="white",names=c("CUSUM","SHEWHART","COMB"))
vioplot(ponto.det[,1],ponto.det[,3],col="white",names
=c("CUSUM","COMB"))

#apenas shewhart
ponto.det.s<-matrix(c(0),ncol=1,nrow=r)
for (i in 1:r) {
y<-rbinom(N,size=n,prob=p1.val)
```

```
ponto.det.s[i,1]<-
shew.1(y,n=n,lsccl=lsccl)$ponto.shewhart}
hist(ponto.det.s)
summary(ponto.det.s)
sd(ponto.det.s,na.rm=T)

#apenas cusum
ponto.det.c<-matrix(c(0),ncol=1,nrow=r)
for (i in 1:r) {
y<-rbinom(N,size=n,prob=pl.val)
ponto.det.c[i,1]<-fl.2(y,k=k.val,h=h.c)$ponto}
hist(ponto.det.c)
summary(ponto.det.c)
sd(ponto.det.c,na.rm=T)

#armazena em planilhas
write.table(ponto.det.c,file="caminho/nomearquivo.csv
",sep=";",dec=".",col.names = NA, qmethod = "double")
write.table(ponto.det.s,file="caminho/nomearquivo.csv
",sep=";",dec=".",col.names = NA, qmethod = "double")
write.table(ponto.det,file="caminho/nomearquivo.csv",s
ep=";",dec=".",col.names = NA, qmethod = "double")
```

APÊNDICE F - Rotina no R para construção do gráfico Combinado Shewhart-CUSUM binomial unilateral superior

```

combbinom3.1=function(y,n,k,h,u,lsc){
size=rep(n,length(y))
if (h<=0)
stop ("h deve ser positivo")
if (u < 0)
stop ("valor inválido para u")
if (lsc <= 0)
stop ("valor inválido para LSC")
zn=u
zn[0]=u
for (n in 1:length(y)){zn[n]=max(0,zn[n-1]+y[n]-k)}
pontos.fora.cusum=subset(zn,zn>h)
pontos.fora.shewhart=subset(y,y>lsc)
plot(y,type='b',ylim=c(0,max(1.01*zn,1.01*h)),pch=21,
bg="gray90", xlab="Amostras",ylab="itens não
conformes")
abline(h=h,lty=1)
#text(length(zn)-2,h+0.6,paste("h=",h))
abline(h=lsc,lty=2)
#text(length(zn)-2,lsc+0.6,paste("LSC=",lsc))
lines(zn,type='b',pch=19)
legend("topleft",
c(expression(paste(C["+"])), "x", "h", "LSC"),
pch=c(19,21,NA,NA),lty=c(1,1,1,2),pt.bg=c("black","gr
ay90", "gray90", "black"))
structure(list(vetor=zn,k=k,h=h,LSC=lsc,pontos.fora.c
usum=length(pontos.fora.cusum),valores.cusum=pontos.f
ora.cusum,pontos.fora.shewhart=length(pontos.fora.she
whart),valores.shewhart=pontos.fora.shewhart))
}

```

APÊNDICE G – Rotina no R para construção do gráfico CUSUM binomial unilateral superior

```
f1=function (kmais,hmais,x){
cmais=0
cmais[0]=0
for (i in 1:length(x)){cmais[i]=max(0,cmais[i-
1]+x[i]-kmais)}
plot(cmais,type='b',ylim=c(0,max(cmais+1,hmais)),ma
in='
',ylab=expression(paste(C["+"])),xlab='Amostras',lw
d=1,bg=c("white"))
points(cmais,pch=19)
abline(h=hmais,lty=1)
legend("topleft", c(expression(paste(C["+"])), "h"),
pch=c(19,NA),lty=c(1,1))
#text(2,hmais+0.5,paste("h=",hmais))
#grid(20)
return(cmais)
}
```

APÊNDICE H – Rotina no R para as aplicações do Capítulo 8

```

library(surveillance)
library(qcc)
library(vcd)
library(adehabitac)

###APLICACAO XICARAS

library(Devore7)

###devore dados exemplo 16.6 pag 620

data(xmp16.06)
?xmp16.06

x1<-xmp16.06[3:27,1:3]
x1
n=100

p0=0.0608
p1=1.2*0.0608
###calculando ARL combinado

arlCOMB.1(370)
arlCOMB.1(370,2852)

###limites
qbinom(1/2852,prob=p0,size=100,lower=F)
1-pbinom(16,prob=p0,size=100)
1/(1-pbinom(15,prob=p0,size=100))
qnorm(1/2852)#3.4

## calculando LSC do SHEWHART
qcc(x1$C2,sizes=100,type="np")

###calculando os parametros do CUSUM

##h
##aproximacao CD (extraindo a raiz)
hrs.1(0.0608,1.2*0.0608,n=100,425)
hrs.5(0.0608,1.2*0.0608,n=100,arl0=425,digits=2,tol=0.05)

```



```

###colocando uma "grande alteracao" em x2 "mantendo a
média" em aproximadamente p1=0,073
x2.1<-c( 4, 3 , 6, 8,17 , 5 , 6 , 7 , 7, 11 , 5 , 3
, 7, 7, 8, 10, 8 , 6, 10 , 2 ,11 , 7 , 8, 10, 7)
mean(x2.1)
x3.1=c(x1$C2,x2.1)

##plotando o gráfico combinado
combinom3.1(x3.1,n=100,k=6.7,h=17.1,u=0,lsc=15)
abline(v=25,lty=2,col="gray",lwd=2)#coloca uma linha
pontilhada vertical apos a amostra 25
text(5,17.5,paste("h=17,1"),cex=1.1)
text(5,15.4,paste("LSC=15"),cex=1.1)

##apenas shewhart
qcc(x3.1,limits=c(0,13.25),sizes=100,center=6.08,type
="np")
abline(v=25,lty=2,col="gray",lwd=2)

#apenas CUSUM
f1(x3.1,kmais=6.7,hmais=16.5)
abline(v=25,lty=2,col="gray",lwd=2)
text(5,17,paste("h=16,5"),cex=1.1)

###APLICACAO COLIFORMES

x=read.csv("C:\\Users\\Elisa\\Desktop\\TESE\\cap8\\da
dosCOLIFORMES.csv",sep=";",dec=",",header=T)
names(x)

###SUPOSICOES

###autocorrelacao
par(mfrow=c(1,2))
acf(x$NC[1:24],xlim=c(1,13),main="ACF")
pacf(x$NC[1:24],xlim=c(1,13),main="PACF")
par(mfrow=c(1,1))
wawotest(x$NC[1:24])

### superdispersao
size=rep(205,length(x$NC[1:24]))
qcc.overdispersion.test(x$NC[1:24],size,type="binomia
l")

```

```

###aderencia
gf <- goodfit(x$NC[1:24], type = "binomial", par =
list(size = 205,prob=0.022))
summary(gf1)

table(x$NC[1:24])
obs=c(1 ,1, 3, 2, 6, 5, 2, 1, 2, 1)
p<-dbinom(0:9,size=205,prob=0.022)
chisq.test(obs, p = p,rescale.p = TRUE)
chisq.test(obs, p = p,rescale.p =
TRUE,simulate.p.value=T)

#####ARL0=370
nmax=0.05*200

###shewhart
qcc(x$NC[1:24],type="p",sizes=205)
media=round(mean(0.0214*205),1)
p0=0.022
p1=0.05
s=0.05/p0
#s=127%
p1=1.6*p0
}

##cusum ARL0=465.5

p1=1.2*p0
k.val=round(findK(p0,p1,n=205,distr=c("binomial")),2)
h.val=round(hrs.1(p0,p1,n=205,arl0=469)$h2,1)
hrs.5(p0,p1,n=205,arl0=469,tol=0.05,digits=2)
arlCusum(k=4.95,h=17.6,n=205,theta=p0,distr=c("binomi
al"),digits=2)

arlCOMB(469,65817)

###gráfico
combinom3.1(y=x$NC,n=205,k=4.95,h=17,lsc=13,u=0)
text(34.4,19,paste("h=18,0"),cex=1.2)
text(35,14,paste("LSC=13,0"),cex=1.2)
abline(v=24,col="gray",lty=2,lwd=2)

###
arlCOMB.1(465.5,469)
qbinom(1/62244,size=205,prob=p0,lower=F)

```

```

1/(1-pbinom(15, size=205, prob=p0))
arlCOMB.1(465.5, 469)
arlCOMB(473, 65817)

###gráfico
combinom3.1(y=x$NC, n=205, k=k.val, h=h.val, lsc=15, u=0)
text(34.4, 18.6, paste("h=17.6"), cex=1.2)
text(35, 16, paste("LSC=15, 0"), cex=1.2)
abline(v=24, col="gray", lty=2, lwd=2)

###shewhart
plot(x$NC, type="b", col=ifelse(x$NC >
11, 2, 1), pch=21, bg=ifelse(x$NC > 11, 2, 4), main="
", xlab="Amostra", ylab="y")
abline(h=11, lty=2)
#abline(h=10, col=4, lty=2)
abline(v=24, col="gray", lty=2, lwd=2)
text(5, 12, paste("LSC=11, 0"), cex=1.2)

qcc(x$NC[1:24], type="np", nsigmas=2.6, sizes=205)

## apenas cusum

p1=0.05

k.val=round(findK(p0, p1, n=205, distr=c("binomial")), 2)
h.val=round(hrs(p0, p1, n=205, arl0=465.5)$root, 1)
arlCusum(k=6.9, h=5.21, n=205, theta=p0, distr=c("binomia
l"), digits=2)
f1(kmais=6.9, h=5.21, x$NC)
text(34, 6.21, paste("LSC=5.21"), cex=1.2)
abline(v=24, col="gray", lty=2, lwd=2)

colif.totais.1<-x$NC[1:24]
colif.totais.2<-x$NC[25:36]
qcc(colif.totais.1, type="np", limits=c(0, 11), center=20
5*p0, newdata=colif.totais.2, sizes=205, newsizes=205)

#####

###rotulagem - exemplo 3

x=c(3, 4, 1, 2, 4, 1, 2, 1, 2, 4, 0, 1, 2, 0, 4, 1, 1, 1)
qcc(x, type="p", sizes=36)

```

```
p0=0.052
x1=c(5,1,0, 2,4,0, 0,1,9, 2,0,5, 4,1,1, 1,0,1, 9,2,0,
      5,4,1, 1,1,4, 6,6,6, 7,3,4, 3,4,1, 1,1,4, 6,6,6, 7,
      3,4,3, 3,4,2, 8,5,9)
```

```
###autocorrelação
par(mfrow=c(1,2))
acf(x,main=" ",xlim=c(1,12))
pacf(x,main=" ",xlim=c(1,12))
par(mfrow=c(1,1))
```

```
###autocorrelação com dados transformados
y=asin(sqrt((x+3/8)/(36+0.75)))
par(mfrow=c(1,2))
acf(y,main=" ",xlim=c(1,12))
pacf(y,main=" ",xlim=c(1,12))
par(mfrow=c(1,1))
```

```
wawotest(x)
```

```
### superdispersao
size=rep(36,length(x))
qcc.overdispersion.test(x,size,type="binomial")
```

```
###aderencia
gf1 <- goodfit(x, type = "binomial", par = list(size
= 36,prob=p0))
summary(gf1)
```

```
obs=c( 2, 7 ,4 ,1, 4)
p<-dbinom(0:4,size=36,prob=0.052)
chisq.test(obs, p = p,rescale.p =
TRUE,simulate.p.value=T)
chisq.test(obs, p = p,rescale.p =
TRUE,simulate.p.value=T)$expected
```

```
###grafico Shewhart
p0=0.052
qcc(x,type="np",sizes=36)
pbinom(5.9,size=36,prob=p0,lower=F)
1/pbinom(5.9,size=36,prob=p0,lower=F)
1/(1-pbinom(6,size=36,prob=p0))
```

```
arl0=167
```

```

np1=36*p0+0.5*sqrt(p0*(1-p0)*36)
p1=np1/36
p1=0.07
s=p1/p0-1

hrs.l(p0,p1,n=36,548)
arlCusum(k=2.18,h=11.25,n=36,distr=c("binomial"),theta
a=p0,digits=2)

arlCOMB(556,2298.4)
qbinom(0.0001916165,size=36,prob=p0,lower=F)
1/pbinom(7,size=36,prob=p0,lower=F)
arlCOMB.l(443,2298.4)

x2=c(x,x1)
lsc=7
h=11.3
k=2.18

###combinado
combbinom3.1(y=x2,n=36,k=2.18,h=11.3,u=0,lsc=7)
abline(v=18,col="gray",lwd=2,lty=2)
text(65,12.3,paste("h=11,3"),cex=1.1)
text(65,8.2,paste("LSC=7"),cex=1.1)

###cusum individual
#mesmo p1 e arl0
hrs.l(p0,p1,n=36,443)
arlCusum(k=2.18,h=10.6,n=36,distr=c("binomial"),theta
=p0,digits=2)
f1(kmais=2.18,hmais=10.6,x=x2)
abline(v=18,col="gray",lwd=2,lty=2)
text(65,12,paste("h=10,6"),cex=1.1)

#p1 maior
p1=2*p0
hrs.l(p0,p1,n=36,443)
arlCusum(k=2.7,h=5.9,n=36,digits=2,theta=p0,distr=c("
binomial"))
f1(kmais=2.7,hmais=5.9,x2)
abline(v=18,col="gray",lwd=2,lty=2)
text(65,7,paste("h=5,9"),cex=1.1)

###shewart sem regras suplementares

```

```
plot(x2,type="b",pch=19,col=ifelse(x2>6,2,1),main="
",xlab="Amostra",ylab="np")
abline(h=6,col=2,lwd=2)
text(5,6.2,paste("LSC=6"),cex=1.1)

###shewart com regras suplementares
qcc(x2,type="np",sizes=36,center=36*p0,limits=c(0,6))
```

APÊNDICE I – Verificação da aderência

A verificação da aderência à distribuição binomial foi complementada mediante um teste de adEquação de ajuste baseado na razão de verossimilhança do pacote **vcd** (MEYER; ZEILEIS; HORNIK, 2009). A função `goodfit` aproxima valores para uma distribuição discreta (Poisson, binomial ou binomial negativa) para dados de contagens. Dois métodos podem ser escolhidos: Máxima Verossimilhança e Qui Quadrado mínimo. Os valores correspondentes as estatísticas qui quadrado obtidas pelos dois métodos respectivos p-valores são retornados. Assim pode ser feita uma comparação se os dois conjuntos de valores, observados e ajustados, pertencem a mesma distribuição. A hipótese H_0 é a de que os dois conjuntos de dados pertencem à mesma distribuição. A hipótese H_1 é a de que os dois conjuntos de dados não pertencem à mesma distribuição. A Tabela I.1 traz os valores esperados (obtidos pelo ajuste) para os dados exemplo de aplicação 8.2 (coliformes totais). Os valores de X (primeira coluna) correspondem às contagens. A estatística qui quadrado obtida pela razão da máxima verossimilhança é 5,875 e o p-valor = 0,752, com grau de liberdade igual a 9. Na Figura I.1 podem ser comparados os valores esperados e observados.

Tabela I. 1: Valores observados e esperados para os dados do exemplo 8.2.

X	Observados	Ajustados
0	1	0,25
1	1	1,16
2	3	2,66
3	2	4,04
4	6	4,59
5	5	4,15
6	2	3,11
7	1	1,99
8	2	1,10
9	1	0,54

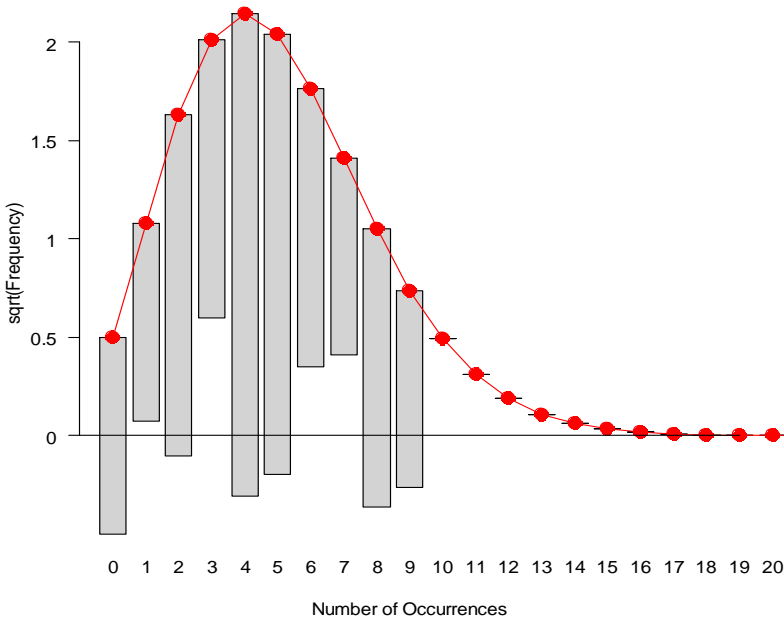


Figura I. 1 - Teste de aderência para dados do Exemplo 8.2

A Tabela F.2 traz os valores esperados (obtidos pelo ajuste) para os dados exemplo de aplicação 8.3 (rotulagem de garrafas). A estatística qui quadrado obtida pela razão da máxima verossimilhança é 7,41 e o p-valor = 0,116, com grau de liberdade igual a 4. Na Figura I.2 podem ser comparados os valores esperados e observados.

Tabela I. 2: Valores observados e esperados para os dados do exemplo 8.3.

X	Observados	Ajustados
0	2	2,63
1	7	5,20
2	4	4,99
3	1	3,10
4	4	1,40

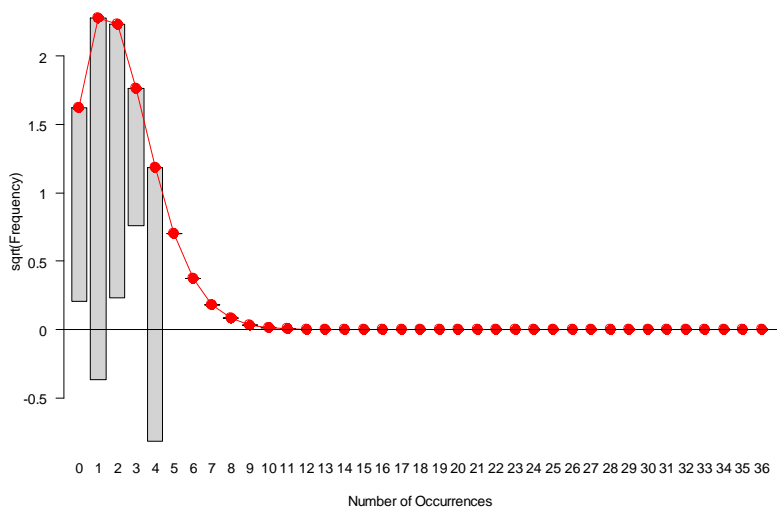


Figura I. 2 - Teste de aderência para dados do Exemplo 8.3.

Para ambos os conjuntos de dados pode-se aceitar a suposição de aderência à distribuição binomial. O nível de significância adotado foi de 5%.

ANEXO A – Dedução do valor de referência k do SPRT

A estatística usada pelo SPRT (Sequential Probability Ratio Test) é definida em termos de um teste de razão de máxima verossimilhança das densidades das variáveis aleatórias envolvidas. Para a distribuição de Bernoulli a função de probabilidades de X_i é $f(x, p) = p^x(1 - p)^{1-x}$. Após a inspeção de j itens, essa estatística fica,

$$\Lambda_j = \sum_{i=1}^j L_i; \quad (\text{A.1})$$

Onde,

$$L_i = \ln \frac{f(X_i; p_1)}{f(X_i; p_0)};$$

$$L_i = \ln(p_1^{X_i}(1 - p_1)^{1-X_i}) - \ln(p_0^{X_i}(1 - p_0)^{1-X_i});$$

$$L_i = \ln p_1^{X_i} + \ln(1 - p_1)^{1-X_i} - \ln p_0^{X_i} - \ln(1 - p_0)^{1-X_i};$$

$$L_i = X_i \ln p_1 + (1 - X_i) \ln(1 - p_1) - X_i \ln p_0 - (1 - X_i) \ln(1 - p_0);$$

$$L_i = X_i(\ln p_1 - \ln(1 - p_1) - \ln p_0 + \ln(1 - p_0)) + \ln(1 - p_1) - \ln(1 - p_0);$$

portanto,

$$L_i = X_i \left\{ \ln \left(\frac{p_1(1-p_0)}{p_0(1-p_1)} \right) \right\} + \ln \left(\frac{1-p_1}{1-p_0} \right). \quad (\text{A.2})$$

As quantidades nos logaritmos são chamadas, para simplificar, de r_2 e $(-r_1)$, respectivamente, ficando então;

$$L_i = X_i r_2 - r_1; \quad (\text{A.3})$$

onde substituindo na Equação (3.10) fica-se com

$$\Lambda_j = T_j r_2 - r_1 j. \quad (\text{A.4})$$

Com $\Lambda_j = \sum_{i=1}^j X_i$ que corresponde ao número total de itens defeituosos nos j itens inspecionados. O SPRT requer a especificação de duas constantes A e B , com $A < B$, e uma regra de decisão.

A desigualdade $A < \Lambda_j < B$ determina quando o SPRT deve continuar o processo de amostragem. Os autores complementam que para algumas aplicações é mais conveniente substituir $T_j r_2 - r_1 j$ em $A < \Lambda_j < B$ e dividir cada termo por r_2 . Assim, tem-se

$$\frac{A}{r_2} < T_j - \gamma j < \frac{B}{r_2}. \quad (\text{A.5})$$

Complementando, a regra de decisão é

- i. Se $T_j - \gamma j \geq \frac{B}{r_2}$ interrompe a amostragem e rejeita H_0 ;
- ii. Se $T_j - \gamma j \leq \frac{A}{r_2}$ interrompe a amostragem e aceita H_0 .

Onde,

$$\gamma = \frac{r_1}{r_2} = \frac{\ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}. \quad (\text{A.6})$$

A expressão A.6 deduzida para Bernoulli, com $n = 1$. Estendendo para a distribuição binomial, tem-se $k = n\gamma$.

ANEXO B – Valor de referência k para as distribuições Normal e de Poisson

Neste anexo, verifica-se como fica k, para algumas distribuições conhecidas. Para o caso da distribuição normal, esta pode ser escrita na forma da família exponencial como

$$f(y|\mu) = e^{\left(\frac{y\mu}{\sigma^2} - \frac{y^2}{2\sigma^2} - \ln(\sigma\sqrt{2\pi}) - \frac{\mu\mu^2}{\sigma^2}\right)}. \quad (\text{B.1})$$

E, deste modo, tem-se: $a(y) = y$, $b(\mu) = \frac{\mu}{\sigma^2}$ e $d(\mu) = -\frac{\mu^2}{2\sigma^2}$.

Substituindo-se na expressão para k, fica-se com

$$k = -\frac{\frac{(-\mu_1^2 + \mu_0^2)}{2\sigma^2}}{\frac{\mu_1 - \mu_0}{\sigma^2}} = \frac{\mu_0 + \mu_1}{2}. \quad (\text{B.2})$$

Para a distribuição de Poisson, o parâmetro k é sumarizado por

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} = \frac{\mu_1 - \mu_0}{\ln\left(\frac{\mu_1}{\mu_0}\right)}, \quad (\text{B.3})$$

onde

$$f(y|\lambda) = e^{y \ln \mu - \mu - \ln \Gamma(\mu+1)}, \quad (\text{B.4})$$

com $d(\theta) = d(\mu) = -\mu$ e $b(\theta) = b(\mu) = \ln \mu$.

ANEXO C – Aproximação de Siegmund

Suponha um processo capaz de produzir uma seqüência infinita de observações x_1, x_2, \dots . Inicialmente este processo é considerado “sob controle” estatístico no sentido de que uma observação é suficiente para guardar as observações sem tomar nenhuma ação. Em algum momento desconhecido v o processo muda e se torna “fora de controle” estatístico. O pesquisador gostaria de inferir em que ponto esta mudança se iniciou e tomar uma ação apropriada “o quanto antes possível” após o momento v .

Assume-se que a variável x é uma variável aleatória independente e que para $v \geq 1$, x_1, x_2, \dots, x_{v-1} tem função densidade de probabilidade f_0 , ao passo que x_v, x_{v+1}, \dots tem densidade de probabilidade f_1 . Denota-se por P_v a probabilidade da mudança de f_0 para f_1 ocorrer na v -ésima observação, $v = 1, 2, \dots$; e faça P_0 ser a probabilidade de não ocorrer mudança. Então, x_1, x_2, \dots são independentes e identicamente distribuídas com função densidade de probabilidade f_0 . Procura-se uma regra de decisão τ que faça a distribuição de P_v de $(\tau - v)^+$ estocasticamente pequena e a distribuição de P_0 de τ seja estocasticamente grande. Uma simples, mas formal condição, minimizar a quantidade $\sup_{v \geq 1} E_v(\tau - v + 1 | \tau \geq v)$.

Uma proposição para a solução aproximada deste problema apresentada por Siegmund (1985) é considerar $1 \leq v \leq n$, sob a hipótese H_v , x_1, \dots, x_{v-1} distribuídas de acordo com f_0 , e x_v, \dots, x_n sob f_1 , e H_0 a hipótese de não haver mudança. O log da razão de verossimilhança para testar H_v contra H_0 é $\sum_{k=v}^n \log(f_1(x_k)/f_0(x_k))$. Para testar esta composição de hipóteses considere que para pelo menos uma das hipóteses H_v , o log da razão de máxima verossimilhança é

$$\max_{0 \leq k \leq n} (\tilde{S}_n - \tilde{S}_k) = \tilde{S}_n - \min_{0 \leq k \leq n} \tilde{S}_k, \quad (C.1)$$

onde $\tilde{S}_n = \sum_{j=1}^n \log[f_1(x_j)/f_0(x_j)]$. Uma regra de decisão intuitiva atraente baseada na Equação anterior é

$$\tau = \inf\{n: \tilde{S}_n - \min_{0 \leq j \leq n} \tilde{S}_j \geq b\}. \quad (C.2)$$

Pode ser notado que $\tilde{S}_n - \min_{0 \leq j \leq n} \tilde{S}_j$ mede o a altura corrente do passeio aleatório $\tilde{S}_j, j = 0, 1, \dots$ sobre o valor mínimo. Sempre que o passeio aleatório estabelece um novo mínimo, o processo ignora o passado e começa novamente considerando que para todo $j \geq 0, \tilde{S}_{n+j} - \min_{0 \leq k \leq n+j} \tilde{S}_k = \tilde{S}_{n-j} - \tilde{S}_n - \min_{0 \leq k \leq j} (\tilde{S}_{n+k} - \tilde{S}_{n_j})$.

Esta renovação, proposta por Siegmund (1979, 1985), tem conseqüências importantes. Primeiro implica que para cada τ definido, $\sup_{v \geq 0} E_v(\tau - v + 1 | \tau \geq v) = E_1 \tau$, pois em todo tempo após $v - 1$ o processo necessita ter no mínimo a amplitude da renovação naquele instante. Por isso para calcular as

Equações anteriores é preciso determinar $E_\nu(\tau)$ para $\nu = 0,1$, e, para cada um destes casos “extremos”, x_1, x_2, \dots são identicamente distribuídas. Isto significa que a regra de decisão τ pode ser definida em termos de uma seqüência de testes de razão de probabilidades, como mostrado a seguir. Siegmund (1985) denota que

$$N_1 = \inf\{n: \tilde{S}_n \notin (0, b)\} \quad (C.3)$$

Se $\tilde{S}_{N_1} \geq b$, então $\tau = N_1$. Caso contrário $\tilde{S}_{N_1} \leq \min_{0 \leq k \leq N_1} \tilde{S}_k$, e definindo-se

$$N_2 = \inf\{n: n \geq 1, \tilde{S}_{N_1+n} - \tilde{S}_{N_1} \notin (0, b)\}. \quad (C.4)$$

Se $\tilde{S}_{N_1+N_2} - \tilde{S}_{N_1} \geq b$ então $\tau = N_1 + N_2$. Caso contrário, $\tilde{S}_{N_1+N_2} \leq \tilde{S}_{N_1}$, e $\tilde{S}_{N_1+N_2} = \min_{0 \leq k \leq N_1+N_2} \tilde{S}_k$. E, de modo geral

$$N_k = \inf\{n: n \geq 1, \tilde{S}_{N_1+\dots+N_k} - \tilde{S}_{N_1+\dots+N_{k-1}} \notin (0, b)\}. \quad (C.5)$$

Para o autor, é fácil de ver que

$$\tau = N_1 + \dots + N_M, \quad (C.6)$$

onde

$$M = \inf\{k: \tilde{S}_{N_1+\dots+N_k} - \tilde{S}_{N_1+\dots+N_{k-1}} \geq b\}. \quad (C.7)$$

O exposto anteriormente tem como consequência que sobre qualquer probabilidade P , para variáveis observadas independentes e identicamente distribuídas, em particular para $P = P_1$ ou P_0 , a identidade de Wald (1947) fornece

$$E_\tau = EN_1EM. \quad (C.8)$$

Além disso, para Siegmund (1985) M é uma variável geometricamente distribuída, com valor esperado igual a

$$E(M) = \frac{1}{P(\tilde{S}_{N_1} \geq b)}. \quad (C.9)$$

Como resultado o autor obteve a seguinte identidade básica

$$E(\tau) = \frac{E(N_1)}{P(\tilde{S}_{N_1} \geq b)}. \quad (C.10)$$

Essa identidade (A.10) é expressa em termos do valor esperado de tamanho de amostra e da probabilidade de erro de um único teste SPRT individual ter um limite inferior (log) $a = 0$. Na Expressão (C.10) aplicando a identidade de Wald quando $a = 0$, conduz a indeterminação matemática $0/0$. Ainda assim, para Siegmund (1985), é possível aproximar o cálculo, considerando para o lado direito da Equação um valor a negativo arbitrário e fazendo $a \rightarrow 0$. Como resultado derivam-se as seguintes expressões

$$E_0(\tau) = \frac{|e^b - b - 1|}{|\mu_0|} \tag{C.11}$$

e

$$E_1(\tau) = \frac{(e^{-b} + b - 1)}{\mu_1}, \tag{C.12}$$

onde

$$\mu_i = \int \log \frac{f_1(x)}{f_0(x)} f_i(x) dx, \quad i = (0,1). \tag{C.13}$$

Então para o caso de variáveis aleatórias membros da família da distribuição exponencial de um parâmetro, foi possível obter aproximações para o valor de ARL de um CUSUM. As observações devem ser independentes e identicamente distribuídas de acordo com uma pré-fixada distribuição pertencente a família exponencial.

Denota-se por $z(x) = \log f_1(x)/f_0(x)$ e assume-se que P_θ é uma distribuição de $z_n = z(x_n)$ escrita na forma da família exponencial com um parâmetro $P_\theta(z_n \in dx) = dF_\theta(x) = e^{\theta z(x) - \psi(\theta)} dF(x)$. Siegmund (1985) recomenda a padronização desta para obter $\int z(x) dF(x) = 0$ assim, $\psi(0) = 0$. Isto pode ser acompanhado de uma mudança nos eixos de θ , fazendo estes serem pares conjugados de valores de θ , com $\theta_0 \leq \theta \leq \theta_1$ e $\psi(\theta_0) = \psi(\theta_1)$, onde $\psi(\theta)$ corresponde a função geradora dos cumulantes, de tal modo que $\psi(0) = \psi'(0) = 0$ e $\psi''(0) = 1$. Wu (2005) estabelece que $F(x)$ (ou $F_0(x)$) deve também ser uma variável “fortemente” não-aritmética de modo que

$$\lim_{|\lambda| \rightarrow 0} \sup \left| \int_{-\infty}^{+\infty} e^{i\lambda x} dF(x) \right| < 1. \tag{C.14}$$

Para continuar supõe-se então que $\theta \neq 0$ e $\theta < 0$. Isto será denotado por θ_0 e seu conjugado θ_1 . Caso contrário, tem-se θ_1 e seu conjugado θ_0 . A regra de decisão pode ser escrita por $\tau = \inf \{n: \Delta \tilde{S}_n - \min_{0 \leq k \leq n} \Delta \tilde{S}_k \geq b'\}$, onde $\Delta = \theta_1 - \theta_0$ e $b' = \Delta b$. Pode-se então escrever o valor esperado E_{θ_i}

$$E_{\theta_i}(\tau) \cong \frac{e^{[(-1)^i \Delta b] - (-1)^i \Delta b - 1}}{\Delta |\psi'(\theta_i)|}, i = 0, 1. \quad (\text{C.15})$$

Aplicando o limite, fazendo $\theta_i \rightarrow 0$, resulta em

$$E_0(\tau) \cong b^2 \psi''(0). \quad (\text{C.16})$$

Siegmund (1985) ilustra que para o caso especial onde $f_i(x) = \phi\left(x + \frac{(-1)^i}{2}\right)$ e $dF(x) = \phi(x)dx$ com $z(x) = x$, e a distribuição normal (escrita na forma da família exponencial) com média θ e variância igual a Um (1), a Equação(C.16) fica

$$E_{\theta}(\tau) \cong \frac{|e^{(-2\theta b) + 2\theta b - 1}|}{2\theta^2}. \quad (\text{C.17})$$

A expressão permite aproximar o ARL. Todavia não obteve resultados precisos quando comparados com os obtidos numericamente por Van Dobben de Bruyn (1968). Siegmund (1985) aprimora esta Equação, corrigindo o “excesso” da estatística SPRT sobre os limites de aceitação ou rejeição. Ele aproxima o limite do valor esperado deste excesso. Assim, inclui-se na Equação (C.17) uma aproximação que faz com que $e(b) = e(b + 2 * 0,583) = e(b + 1,166)$. Para facilitar a compreensão será feita uma modificação na nomenclatura utilizada daqui para diante. O limite de decisão será chamado de h e o limite corrigido de b . Portanto, a expressão A.17 é sumarizada em A.18, na forma conhecida na literatura como “Aproximação de Siegmund”.

$$ARL \cong E_{\theta}(\tau) \cong \frac{e^{-2\theta(h+1,166)} + 2\theta(h+1,166) - 1}{2\theta^2}. \quad (\text{C.18})$$

Na seção seguinte esta correção será explicada com maiores detalhes.

C.1 CORREÇÃO DO LIMITE DE DECISÃO

Pode ser notado que na Expressão A.18 o limite de decisão h foi corrigido por $b = h + \rho_+ - \rho_-$, onde ρ corresponde a aproximação do excesso sobre o intervalo de decisão h do CUSUM. Siegmund (1985) e Wu (2005) colocam que $\rho_{\pm} = E_0 R_{\pm\infty}$ e

$$\rho^{(a)}(\theta) = E_{\theta}(R_{\infty}^{(a)}) = \frac{E_{\theta}(S_{\tau_+}^{a+1})}{(a+1)E_{\theta}(S_{\tau_+}^1)} \text{ para } a > 0, \quad (\text{C.19})$$

onde R_∞ é uma variável aleatória da distribuição limitante para o resíduo $S_{\tau(b)} - b$, e $\rho^{(a)}$ corresponde aos (a) momentos dessa distribuição. Portanto, o momento de interesse ρ_+ será dado por

$$\rho_+ = \frac{E_0(S_{\tau_+}^2)}{2E_0(S_{\tau_+})}. \tag{C.20}$$

Siegmund (1985) apresenta uma aproximação para ρ_+ , que pode ser facilmente resolvida por métodos numéricos computacionais. Assumindo que a distribuição de z_i é contínua e fortemente não aritmética então ρ_+ pode ser aproximado por

$$\rho_+ = \frac{\kappa}{6} - \int_0^\infty \frac{1}{\pi\lambda^2} Re \left(\log \left(2 \left(\frac{1-g(\lambda)}{\lambda^2} \right) \right) \right), \tag{C.21}$$

onde $g(\lambda) = e^{\psi(i\lambda)}$ e $\kappa = \psi^{(3)}(0)$ corresponde ao terceiro cumulante de $F(x)$.

Siegmund (1985) define que a hipótese de que z tenha uma distribuição contínua é apenas para garantir a convergência de $S_{\tau_+}^{\sigma\sigma} \rightarrow S_{\tau_+}$, podendo esta condição ser minimizada ou eliminada totalmente. Isto ocorre com a condição de ser uma variável não aritmética.

Wu (2005) mostra algumas identidades que utilizam os resultados de Siegmund (1979), formalizados por Chang (1992). Duas dessas identidades estão sumarizadas a seguir.

$$E_0 S_{\tau_+} E_0 S_{\tau_-} = -\frac{1}{2} \tag{C.22}$$

e

$$\rho_+ + \rho_- = \frac{\gamma}{3}. \tag{C.23}$$

Nas identidades A.22 e A.23 κ e γ correspondem ao terceiro e quarto cumulantes de $F(x)$, quando $\theta = 0$. Assim, para distribuição simétrica tem-se $\rho_+ = \rho_-$.

Exemplo C.1: Neste exemplo será aproximado o valor da correção do excesso para uma variável com distribuição normal. Seja X uma variável aleatória contínua com função densidade de probabilidades dada por:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}, \mu \in \mathfrak{R}. \tag{C.24}$$

A função $f(x; \mu, \sigma^2)$ é escrita na forma padrão da família exponencial com um parâmetro $dF_\theta(x) = e^{\theta_z(x) - \psi(\theta)} dF(x)$.

$$f(x; \mu, \sigma^2) = \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right];$$

$$f(x; \mu, \sigma^2) = \exp \left\{ \frac{1}{\sigma^2} \left[x\mu - \frac{\mu^2}{2} \right] - \frac{1}{2} \ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} \right\}.$$

Comparado com a forma padrão tem-se: $\theta = \mu$ e $\psi(\theta) = \frac{\theta^2}{2}$.

Então, $\psi(\theta = 0) = 0$, $\psi^{(\theta)} = \theta$ e $\psi'(\theta = 0) = 0$, $\psi''(\theta = 0) = 1$, $\psi^{(3)}(0) = \gamma = 0$ e $\psi^{(4)}(0) = \kappa = 0$.

E, fazendo $g(\lambda) = e^{(i\lambda)^2/2}$, assim

$$\rho_+ = 0 - \int_0^\infty \frac{1}{\pi\lambda^2} \operatorname{Re} \left(\log \left(2 \left(\frac{1 - e^{(i\lambda)^2/2}}{\lambda^2} \right) \right) \right) d\lambda.$$

Extraindo a parte real da integral fica-se com

$$\rho_+ = - \int_0^\infty \left(\frac{\log \left(\frac{2}{\lambda^2} - \frac{2e^{-\lambda^2/2}}{\lambda^2} \right)}{\pi\lambda^2} \right) d\lambda.$$

Resolvendo a integral²²: $\rho_+ = 0 - (-0,583) = 0,583$. Como para a normal $\rho_+ + \rho_- = 0$, $\rho_- = -\rho_+$ então $\rho_- = -0,583$. Substituindo o valor de ρ na expressão $b = h + \rho_+ - \rho_- = h + 0,583 - 0,583 = h + 2 * 0,583 = h + 1,16$, onde h corresponde ao limite de decisão do CUSUM.

Woodall e Adams (1993) e Reynolds e Stoumbos (1998, 1999) utilizaram a aproximação de Siegmund no planejamento de gráficos CUSUM e SPRT para o monitoramento da média de uma característica da qualidade com distribuição normal.

²²Função *integrate* (rotinas quadpack) do GNU R.

ANEXO D -Função findH do R

```

##Rotina da função findH do pacote Surveillance
(HÖHLE, 2007) para calcular o limite superior h

function (ARL0, theta0, s = 1, rel.tol = 0.03, roundK
= TRUE, distr = c("poisson", "binomial"), digits = 1,
FIR = FALSE, ...)
{distr <- match.arg(distr, c("poisson", "binomial"))
  fir.arl <- ifelse(FIR, 2, 1)
  thetal <- getThetal(theta0, s = s, distr = distr)
  k <- findK(theta0, thetal, roundK = roundK, distr
= distr, digits = digits, ...)
h1 <- min(12, 4 * k)
  arl1 <- arlCusum(h = h1, k = k, theta = theta0,
distr = distr, digits = digits, ...)[[fir.arl]]
  nEval <- 1
  while (arl1 < 100) {
    h1 <- 2 * h1
    arl1 <- arlCusum(h = h1, k = k, theta =
theta0, distr = distr,
      digits = digits, ...)[[fir.arl]]
    nEval <- nEval + 1}
  h2 <- h1 * 2^(sign(ARL0 - arl1))
  arl2 <- arlCusum(h = h2, k = k, theta = theta0,
distr = distr, digits = digits, ...)[[fir.arl]]
  nEval <- nEval + 1
  h.hat <- round(logInterpolation(ARL0, h1, h2,
arl1, arl2),
    digits)
  arl <- arlCusum(h = h.hat, k = k, theta = theta0,
distr = distr,
    digits = digits, ...)[[fir.arl]]
  nEval <- nEval + 1
  relTol <- abs((arl - ARL0)/ARL0)
  i <- 0
  signs <- sign(ARL0 - arl)
  convergence <- relTol < rel.tol
  if (convergence) {
    return(c(theta0 = theta0, h = h.hat, k = k,
ARL = arl,
      rel.tol = relTol))}
  denrat <- 1/commonDenom(1, k, digits = digits)
  steps <- denrat

```

```

    hLow <- round(h.hat + signs * steps, digits)
    arlLow <- arlCusum(h = hLow, k = k, theta =
theta0, distr = distr, digits = digits,
...) [[fir.arl]]
    nEval <- nEval + 1
    relTol.Low <- abs((arlLow - ARL0)/ARL0)
    if (relTol.Low < rel.tol) {
        return(c(theta0 = theta0, h = hLow, k = k,
ARL = arlLow,
            rel.tol = relTol.Low))}
    while (sign(ARL0 - arl) * sign(ARL0 - arlLow) >
0) {
h.hat <- hLow
        arl <- arlLow
        relTol <- relTol.Low
        signs <- sign(ARL0 - arl)
        hLow <- round(h.hat + signs * steps, digits)
        arlLow <- arlCusum(h = hLow, k = k, theta =
theta0, distr = distr,
            digits = digits, ...) [[fir.arl]]
        nEval <- nEval + 1
        relTol.Low <- abs((arlLow - ARL0)/ARL0)
        if (relTol.Low < rel.tol) {
            return(c(theta0 = theta0, h = hLow, k =
k, ARL = arlLow,
                rel.tol = relTol.Low))}
    if (sign(ARL0 - arlLow) < 0) {
        h.hat <- hLow
        arl <- arlLow
        relTol <- relTol.Low}
    return(c(theta0 = theta0, h = h.hat, k = k, ARL =
arl, rel.tol = relTol)).

```

ANEXO E – Gráficos violin plots

Violin plots (ou apenas *vioplot*) é um método gráfico para análise de dados quantitativos. Este gráfico é uma combinação do boxplot e de um gráfico de densidade kernel. Especificamente ele começa com o boxplot. Então é acrescentado um gráfico de densidade *Kernel* rotacionado, em cada lado do boxplot. Este gráfico é similar ao boxplot, com exceção que também mostra a densidade de probabilidade dos dados. Nestes gráficos há indicação da mediana e da distancia interquartil. Sobreposto ao boxplot está uma estimativa de densidade *Kernel*. Sua forma é semelhante à de um violino, daí o nome recebido (HINTZE, 1998; NATIONAL INSTITUTE OF STANDARDS NA TECHNOLOGY, 2003).

Violin plots são encontrados em extensões de alguns pacotes estatísticos e no GNU R, em alguns pacotes, como o *vioplot* (ADLER, 2005).

Na Figura E.1 estão três exemplos de gráficos violin plot gerados com auxílio do pacote *vioplot* do R, a partir de 2000 dados simulados com distribuição bimodal, uniforme e normal (ADLER, 2005). A mediana é representada pelo ponto branco. A Figura E.2 tem gráficos boxplot para os mesmos dados simulados. Através dos destes gráficos é possível visualizar melhor a distribuição dos dados, como no caso das duas modas e a uniformidade. É, numa situação bem simples, como reunir um histograma ao boxplot.

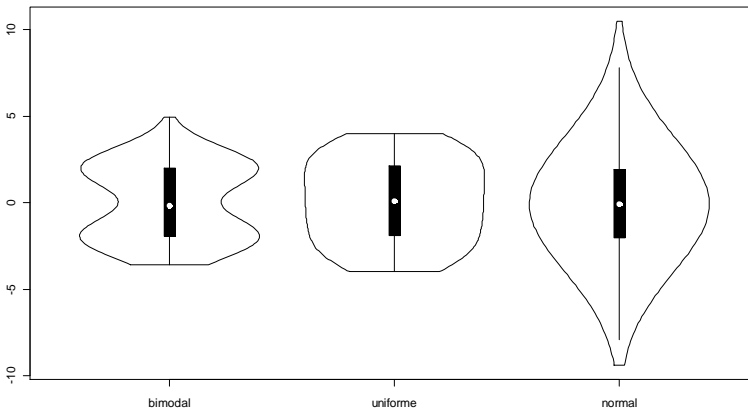


Figura E. 1: Violin plots para dados simulados com distribuição bimodal ($\mu_1=-2$, $\mu_2=2$ e $\sigma_1=\sigma_2=0,6$), uniforme ($\alpha=-4$ e $\beta=4$) e normal ($\mu=0$ e $\sigma=3$).

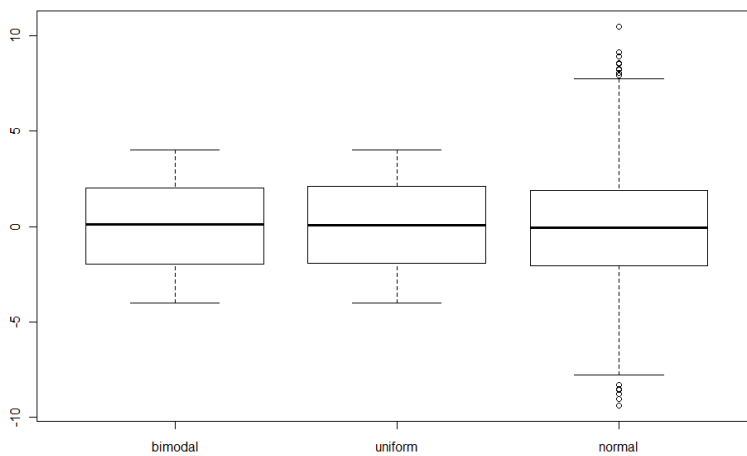


Figura E. 2 - Boxplots para dados simulados com distribuição bimodal ($\mu_1 = -2, \mu_2 = 2$ e $\sigma_1 = \sigma_2 = 0,6$), uniforme ($\alpha = -4$ e $\beta = 4$) e normal ($\mu = 0$ e $\sigma = 3$).

**ANEXO F – Conjunto de dados do exemplo de aplicação 8.2:
Número de amostras não conformes (com presença de coliformes
totais)**

**Tabela F. 1 - Conjunto de dados do exemplo de aplicação 8.2: Número
de amostras não conformes (com presença de coliformes totais)**

Ano	Mês	NC	Ano	Mês	NC	Ano	Mês	NC
2007	jan	8	2008	jan	6	2009	jan	3
2007	fev	4	2008	fev	4	2009	fev	5
2007	mar	8	2008	mar	5	2009	mar	8
2007	abr	2	2008	abr	5	2009	abr	12
2007	mai	2	2008	mai	5	2009	mai	7
2007	jun	2	2008	jun	3	2009	jun	7
2007	jul	0	2008	jul	5	2009	jul	19
2007	ago	4	2008	ago	1	2009	ago	6
2007	set	6	2008	set	3	2009	set	8
2007	out	4	2008	out	4	2009	out	8
2007	nov	7	2008	nov	9	2009	nov	6
2007	dez	4	2008	dez	5	2009	dez	7

Fonte: Relatórios Anuais de Qualidade dos Anos de 2007, 2008 e 2009 da Companhia de Água e Saneamento “Águas de Joinville”

OBS: NC corresponde a o número de amostras consideradas não conformes.

ANEXO G – Conjunto de dados do exemplo 8.3: Processo de rotulagem de garrafas

Tabela G 1 - Conjunto de dados do exemplo 8.3: Processo de rotulagem de garrafas.

NA	NC	NA	NC	NA	NC	NA	NC	NA	NC
1	3	15	4	29	0	43	1	57	4
2	4	16	1	30	5	44	1	58	6
3	1	17	1	31	4	45	4	59	6
4	2	18	1	32	1	46	6	60	6
5	4	19	5	33	1	47	6	61	7
6	1	20	1	34	1	48	6	62	3
7	2	21	0	35	0	49	7	63	4
8	1	22	2	36	1	50	3	64	3
9	2	23	4	37	9	51	4	65	3
10	4	24	0	38	2	52	3	66	4
11	0	25	0	39	0	53	4	67	2
12	1	26	1	40	5	54	1	68	8
13	2	27	9	41	4	55	1	69	5
14	0	28	2	42	1	56	1	70	9

Fonte: Dados fornecidos pela empresa.

OBS:

NA é o número da amostra.

NC corresponde a o número de garrafas onde o processo de rotulagem foi considerado não conforme.

ANEXO H – Aderência à distribuição binomial

Testes de aderência (ou adequação ao ajuste) para distribuições discretas não foram tão extensivamente pesquisadas quanto aos direcionados para distribuições contínuas. Um gráfico de probabilidade binomial (BIPP – Binomial Probability Paper) foi descrito por Mosteller e Tukey (1949) e sua aplicação na construção de gráficos de controle é discutida por Duncan (1986).

Para a distribuição binomial dois procedimentos populares são os testes qui quadrado de Pearson (de aderência) e teste de Fisher – Fisher’s binomial index-of-dispersion test (JOHNSON; KEMP; KOTZ, 2005). Discussões sobre a aplicação destes testes são encontrados em Best e Rayner (1997); Bennet e Hsu (1961).

Os testes qui quadrado e de dispersão de Fisher foram considerados mais poderosos que os baseados na distribuição empírica. A razão apontada pelos autores para a pobre sensibilidade destes (distribuição empírica) era sua forma: bilateral para a distribuição empírica e unilateral nos demais Best e Rayner (1997).

Johnson; Kemp e Kotz (2005) listam uma série de trabalhos que discutem variações nestes testes e outros, incluindo a correção de Yates, distribuição assintótica da estatística de verossimilhança e também abordagem bayesiana.

O teste qui quadrado usual (também conhecido como qui quadrado de Pearson) tem deficiências quando os valores esperados das contagens são pequenos, inferiores a cinco por exemplo. A aproximação da estatística qui quadrado nestes casos pode ser feita utilizando a razão da máxima verossimilhança (RMV) (ÖZDEMIR e EYDURAN, 2005). Em termos de poder do teste, Pearson e Razão da Máxima-Verossimilhança produzem resultados similares, com alguma vantagem do segundo (RMV) para valores esperados inferiores a cinco. Best e Rayner (1997) recomendam ainda a aproximação dos p-valores por simulação de Monte Carlo, para valores de contagens pequenos.

Para finalizar, uma adaptação do teste Kolmogorov-Smirnov, para dados discretos, é uma importante alternativa para verificação da aderência à distribuição binomial (BARTHELIS *et al.*, 1978).

