

CONTROLE ESTATÍSTICO DE PROCESSOS
EM MEIOS CONTÍNUOS
(DADOS AUTOCORRELACIONADOS)

Banca Examinadora

Prof. Orientador: Dr. Abraham Laredo Sicsú
Prof.: Dr. Alberto Wunderler Ramos
Prof.: Dr. Wilton de Oliveira Bussab

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO**

ANDRÉ GUILHERME RAMOS WEY

**CONTROLE ESTATÍSTICO DE PROCESSOS EM MEIOS CONTÍNUOS
(DADOS AUTOCORRELACIONADOS)**

Dissertação apresentada ao Curso de Pós-Graduação da FGV/EAESP como requisito para obtenção de título de Mestre em Administração

Área de concentração:
Administração da Produção e
Sistemas de Informação

Orientador: Prof. Dr. Abraham Laredo Sicsú



Fundação Getúlio Vargas
Escola de Administração
de Empresas de São Paulo
Biblioteca



540/2000



1200000540

**SÃO PAULO
2000**

RESUMO

Esta dissertação examina soluções para o controle estatístico em sistemas de produção que utilizam meios contínuos, onde freqüentemente os dados de análise apresentam o fenômeno da autocorrelação.

A princípio são abordadas as técnicas de CEP tradicional (gráficos de Shewhart) para meios discretos, mostrando-se que estas técnicas não dão o resultado esperado quando utilizadas em processos contínuos, tanto em relação à amostragem como ao problema da autocorrelação.

Em seguida são apresentadas outras soluções baseadas em técnicas mais recentes como EWMA e modelos autorregressivos, que tem uma melhor performance em meios contínuos. Damos os aspectos teóricos e exemplos de aplicação de cada solução obtida por consulta bibliográfica.

Finalmente, é feita uma simulação comparativa das soluções citadas, baseando-se em programação de macros e uso de software, para verificar o escopo e desempenho das soluções, sendo dadas recomendações sobre o seu uso, bem como sugestões para futuras pesquisas.

ABSTRACT

This work examines solutions to the statistical control in production systems that use continuous media, where frequently the analysis data present the phenomena of autocorrelation.

In the beginning, the traditional SPC techniques for discrete media (Shewhart's Control Chart) are presented. We show that these techniques do not yield the expected results when used in continuous processes, both in relation to sampling and to the autocorrelation problem.

After that, we present other solutions based on more recent techniques such as EWMA and autoregressive models, which have a better performance in continuous processes. We give the theoretical aspects and application examples of each solution that was obtained by bibliographical research.

Finally, we make a comparative simulation of the mentioned solutions, based on macro programming and use of software, to verify the scope and performance of the solutions, giving recommendations about their use as well as suggestions of future research on the subject.

AGRADECIMENTOS

Ao Prof. Dr. Abraham Laredo Sicsú, meu orientador, pela dedicação e aconselhamento que me foram dispensados ao longo deste trabalho.

Ao Prof. Dr. Alberto Wundeler Ramos, pelo incentivo e sugestões relacionadas ao tema.

Ao Mestre e Doutorando Ricardo Rochman, pela ajuda no uso da linguagem Excel VBA.

A Josane Maria Corciolli, pelo carinho e apoio incondicional nos momentos importantes.

Aos meus pais Luis e Celeste Wey

ÍNDICE

CAPÍTULO 1 – CONCEITOS BÁSICOS DE CONTROLE ESTATÍSTICO DE PROCESSOS	6
1.1-) Controle Estatístico de Processos	6
1.2-) Estudo do Gráfico de Controle	7
1.2.1-) Causas Comuns e Especiais e Definição de Controle Estatístico	7
1.2.2-) Descrição do Gráfico de Controle	11
1.2.3-) Interpretação do Gráfico e Regras de Sensibilidade	14
1.2.4-) Diretrizes para Escolha de Limites de Controle, Tamanho da Amostra e Frequência de Amostragem	15
1.2.5-) Amostragem Racional e Subgrupos Racionais	18
1.3-) Gráfico de Controle para Variáveis	19
1.3.1-) Gráficos de Controle para Média e Amplitude	21
1.3.2-) Gráficos de Controle para Valores Individuais e Amplitude Móvel	29
CAPÍTULO 2 – AUTOCORRELAÇÃO	36
2.1-) Correlação Linear Entre Duas Variáveis	36
2.2-) Autocorrelação	38
2.3-) Influência da Autocorrelação nos Gráficos de Controle Tradicionais	43
CAPÍTULO 3 – SOLUÇÕES DE CEP EM MEIOS CONTÍNUOS COM DADOS AUTOCORRELACIONADOS	48
3.1-) Modelos Autorregressivos (AR(k)) e sua Aplicação ao Controle de Processos Contínuos.	48
3.1.1-) Conceito de Modelos Autorregressivos (AR(k))	48
3.1.2-) Aplicação de Modelos Autorregressivos (AR(k)) ao Controle de Processos	51
3.2-) EWMA (“Exponentially Weighted Moving Average”)	60
3.2.1-) O Gráfico de Controle EWMA	60
3.2.2-) Aplicação do Gráfico EWMA com Dados Autocorrelacionados	61
3.3-) Solução de Wheeler para Dados Autocorrelacionados	69
3.3.1-) Cálculo dos Limites de Controle	69
3.3.2-) Exemplos da Solução de Wheeler	70
CAPÍTULO 4 – ANÁLISE COMPARATIVA DA PERFORMANCE DOS GRÁFICOS DE CONTROLE COM DADOS AUTOCORRELACIONADOS	76
4.1-) Especificação do Teste	76

4.2-) Análise dos Resultados Obtidos	79
4.2.1-) ARL_0	79
4.2.2-) ARL_1 para $k = 1,5$	82
4.2.3-) ARL_1 para $k = 3$	85
4.2.4-) Conclusão	88

BIBLIOGRAFIA	89
---------------------	-----------

APÊNDICE A – TABELAS GERADAS PELA SIMULAÇÃO DA PERFORMANCE DOS GRÁFICOS DE CONTROLE PARA DADOS AUTOCORRELACIONADOS	92
---	-----------

LISTA DE FIGURAS

- Figura 1-1** – Evolução da função densidade de probabilidade para uma variável de controle considerada, mostrando a variação nos parâmetros da distribuição (média e desvio padrão) devidos a causas especiais de variação.
- Figura 1-2** – Gráfico de controle para a média da amostra do nível de iodo no negro de fumo.
- Figura 1-3** – Gráficos de \bar{X} e R para o nível de iodo do negro de fumo.
- Figura 1-4** – Gráficos de X (valores individuais) e RM para a compressão de pelotas.
- Figura 2-1** – Gráfico de dispersão para os pesos e alturas de 33 mulheres.
- Figura 2-2** – Gráfico de dispersão de autocorrelação (lag 1) para o peso molecular de um polímero.
- Figura 2-3** – Gráfico de dispersão de autocorrelação (lag 2) para o peso molecular de um polímero.
- Figura 2-4** – Autocorrelograma para uma variável aleatória (índices de autocorrelação para lags $k=1$ até $k=16$).
- Figura 2-5** – Autocorrelograma para a variável $y(t)$, incluindo os valores discretos para diversos lags.
- Figura 2-6** – Gráficos de valores individuais e amplitude móvel para $y(t)$, variável gerada segundo o modelo AR(1).
- Figura 3-1** – Autocorrelograma para a seqüência de dados do peso molecular de um polímero.
- Figura 3-2** – Gráfico de valores individuais e amplitude móvel (X-RM) para o peso molecular de um polímero (valores observados a cada 2 horas).
- Figura 3-3** – Autocorrelograma dos resíduos do modelo AR(1).
- Figura 3-4** – Gráficos X-RM para os resíduos após aplicação do modelo AR(1).
- Figura 3-5** – Autocorrelograma para a viscosidade de um produto químico.
- Figura 3-6** – Gráficos X-RM para a viscosidade de um produto químico (com a presença de autocorrelação).
- Figura 3-7** – Autocorrelograma para os resíduos do modelo AR(2).
- Figura 3-8** – Gráficos X-RM para os resíduos do modelo AR(2).
- Figura 3-9** - Exemplo de pesos das observações anteriores ao instante t utilizadas no cálculo da estatística EWMA (para $\lambda = 0,15$), em função da idade k da observação.
- Figura 3-10** – Autocorrelograma para as observações de concentração de um produto químico.
- Figura 3-11** – Gráficos X-RM para a concentração de um produto químico.
- Figura 3-12** – Autocorrelograma para os resíduos (erros de previsão) do EWMA com $\lambda = 0,7$.
- Figura 3-13** – Gráficos X-RM para os resíduos (erros de previsão) do EWMA com $\lambda = 0,7$ e limites de Shewhart.
- Figura 3-14** – Autocorrelograma para os resíduos do EWMA no caso do peso molecular de um polímero.
- Figura 3-15** – Gráficos X-RM para os resíduos do EWMA no exemplo do peso molecular de um polímero.
- Figura 3-16** – Gráficos X-RM para a concentração com limites convencionais (não ajustados para autocorrelação).
- Figura 3-17** – Gráficos X-RM com limites ajustados para autocorrelação segundo a solução de Wheeler.
- Figura 3-18** – Gráficos X-RM com limites naturais não ajustados para a autocorrelação ($r_1 = 0,66$).
- Figura 3-19** – Gráficos X-RM para o peso molecular com limites ajustados segundo a solução de Wheeler.

Figura 4-1 – Gráficos comparativos da ARL_0 para as 4 soluções e várias séries geradas, em função da combinação de parâmetros utilizada.

Figura 4-2 – Gráficos comparativos da ARL_1 ($k = 1,5$) para as 4 soluções e séries geradas.

Figura 4-3 – Gráficos comparativos da ARL_1 ($k = 3$) para as 4 soluções e séries geradas.

LISTA DE TABELAS

- Tabela 1-1** - Tipos de gráfico de controle para variáveis em função do tamanho das amostras.
- Tabela 1-2** - Valores coletados para o nível de iodo num carregamento de negro de fumo com $n = 4$; valores de \bar{X} e R já calculados (fonte SCHNEIDER & PRUETT, 1993, pág. 189).
- Tabela 1-3** - Valores coletados para a compressão de pelotas, com amostras de $n = 1$; valores de RM já calculados. (fonte: "CEP Para Processos Contínuos e em Bateladas - Estudos de Caso", Apostila de treinamento do Prof. Alberto Ramos)
- Tabela 2-1** - Pesos (em Kg) e Alturas (em metros) de 33 mulheres.
- Tabela 2-2** - Seqüência de valores de peso molecular de um polímero, acompanhada de séries com os mesmos valores defasados com lag $k=1$ e $k=2$.
- Tabela 2-3** - 100 valores gerados para a variável $y(t)$ de controle de um processo, modelada por uma série AR(1).
- Tabela 2-4** - Número médio de pontos fora dos limites de controle para um modelo AR(1) com $n = 25$.
- Tabela 3-1** - Medidas do peso molecular de um polímero, feitas a cada 2 horas, em função do número da observação t (fonte: MONTGOMERY (1996, pág. 426))
- Tabela 3-2** - Viscosidade de um produto químico. (Fonte: MONTGOMERY, 1996, pág. 428)
- Tabela 3-3** - Valores da concentração de um processo químico. (fonte: MONTGOMERY, 1996, pg. 426)
- Tabela 4-1** - Combinações de parâmetros para os modelos de séries temporais utilizados na simulação.

Capítulo 1 – Conceitos Básicos de Controle Estatístico de Processos

O objetivo deste capítulo é apresentar conceitos estatísticos aplicados aos processos produtivos, ou seja o *Controle Estatístico de Processos*, focalizando uma de suas ferramentas, o *Gráfico de Controle*, bem como estudar os tipos de gráfico que terão maior relevância em nossa dissertação.

O presente capítulo não se destina a ser um texto compreensivo sobre o assunto e sim uma introdução ao CEP, suficiente para nossos objetivos. As principais referências são os trabalhos de MONTGOMERY (1996), WHEELER (1994) e DUNCAN (1985).

Daremos uma idéia intuitiva dos princípios estatísticos nos quais se baseia cada tópico, não nos detendo muito em aspectos teóricos ou demonstrações.

1.1-) Controle Estatístico de Processos

Várias técnicas estatísticas têm sido utilizadas nos processos produtivos e na gestão da qualidade nas organizações. Entre essas técnicas estão o *Controle Estatístico de Processos*, o *Delineamento Experimental* e a *Inspeção por Amostragem*. Nosso trabalho tratará principalmente do *Controle Estatístico de Processos (CEP)* e de uma de suas ferramentas utilizadas para este fim, o *Gráfico de Controle*, desenvolvido nos anos 20 por SHEWHART (1931) na Bell Telephone Laboratories.

Segundo MONTGOMERY (1996), O *Controle Estatístico de Processos (CEP)* é um poderoso conjunto de ferramentas de solução de problemas para se atingir e manter a estabilidade do processo, ou seja, manter constante ao longo do tempo a distribuição da variável ou estatística sendo controlada, e para melhorar a performance do processo pela redução de sua variabilidade.

SHEWHART (apud WHEELER (1994)) dá uma visão do CEP como tendo 3 dimensões para atingir a uniformidade dos produtos:

- **Conceito:** definição do padrão de desempenho ideal no qual o processo gerará produtos tão consistentemente quanto possível.

- **Técnica:** para levar o processo em direção ao seu padrão ideal. Isto se dá com a redução consistente da sua variabilidade (melhoria do processo).
- **Julgamento:** para verificar se o processo chegou ou não próximo ao seu ideal. Este é o monitoramento (“obtenção de raio X”) do processo.

A seguir apresentamos os conceitos básicos do gráfico de controle, iniciando pela definição de causa comum e causa especial de variação de um processo. São abordados os princípios e a base estatística, diretrizes de projeto, normas para a formação de subgrupos e análise de padrões e “regras de sensibilidade” para detectar falta de estabilidade do processo.

1.2-) Estudo do Gráfico de Controle

1.2.1-) Causas Comuns e Especiais e Definição de Controle Estatístico

Para definirmos o que é CEP, tomemos como variável de exemplo o desempenho de um veículo a gasolina (em termos do seu consumo de combustível), medido em km/l toda vez que se enche o tanque do mesmo. Ele não será o mesmo todas as vezes, mas geralmente a variação não será grande desde que certas condições sejam mantidas, como as vias pela qual o veículo trafega e as condições técnicas do mesmo. Vamos supor que nas últimas 25 vezes em que enchemos o tanque o desempenho registrado foi (ler de cima para baixo e da esquerda para a direita):

11,7	11,6	9,8	9,5	11,3
10,4	11,4	11,1	9,8	9,8
10,1	10,2	9,7	11,0	10,9
9,8	10,4	10,2	9,4	9,2
11,1	10,0	9,8	9,0	10,4

Quais serão as causas de variação de um evento para o outro? Elas são difíceis de encontrar e eliminar. Podem ser pequenas regulagens no motor do veículo, imperfeições no asfalto, variações na pressão dos pneus, variações na qualidade do combustível utilizado e muitas outras razões. No entanto, *estas causas não afetam a previsibilidade do processo*. Com um bom grau de acerto, pode-

se afirmar que no futuro o desempenho do veículo será em torno de 10,0 km/l com uma variação de 1,0 km/l para mais ou para menos. Nestas condições diz-se que o processo está sob o efeito de um **sistema estável de causas comuns ou aleatórias**.

Suponhamos agora que nos 5 eventos subsequentes às nossas observações anteriores, o desempenho registrado foi:

10,4 9,1 **7,1** 12,0 9,9

Observemos o valor em **negrito**: aqui nos deparamos com um desvio consideravelmente maior em relação à média dessa variável de controle, que se deve a uma causa de outra natureza, por exemplo sujeira no carburador, tráfego em vias mais congestionadas ou mesmo a troca do motorista por um outro com grande imperícia. A causa que levou a este desvio é chamada de **causa especial ou assinalável** e, diferentemente das causas comuns de variação, sua ação específica pode ser detectada. Como sugestões para evitar que esta causa especial volte a afetar o desempenho pode-se efetuar manutenção preventiva e dar um treinamento específico ao motorista, por exemplo.

Em termos gerais, os processos podem estar somente sob o efeito de causas aleatórias (ou comuns), quando então é possível prever seu comportamento futuro, se bem que não seja possível identificar o pequeno efeito isolado de cada causa aleatória. Quando, além do efeito das causas aleatórias, temos a ação de causas assinaláveis (ou especiais), a previsibilidade do processo é afetada. Estas causas especiais podem e devem ser encontradas e na maioria dos casos eliminadas. Quando eventualmente uma causa especial representar uma melhoria significativa no processo, ela deve ser incorporada ao mesmo.

A forma de se verificar se o processo está sob o efeito de causas comuns apenas ou se há causas especiais é monitorar a distribuição de probabilidade da variável sendo controlada. Quando há apenas causas comuns de variação esta distribuição se mantém ao longo do tempo. A mudança nos parâmetros da distribuição que descreve a variável evidencia a existência de causas especiais de variação.

Tomando como exemplo a figura 1-1 abaixo, a variável “desempenho do veículo” tem uma distribuição com média 10,0 km/l e desvio padrão 1,0 km/l, no instante $t = t_0$. A partir do instante $t = t_1$, a média sofreu um desvio tornando-se 11,0 km/l, mantendo-se o mesmo desvio padrão. Isto evidencia a existência de uma causa especial, por exemplo o uso de combustível aditivado com

melhor desempenho. Esta causa pode ser incorporada ao processo, pois leva a um melhor desempenho. Já, a partir do instante $t = t_2$, o processo continuou na nova média (11,0 km/l), mas o desvio padrão aumentou para 1,5 km/l. Isto também evidencia uma causa especial, como por exemplo uso do veículo em diferentes horários do dia com situações de trânsito distintas, levando a uma maior incerteza quanto ao seu desempenho.

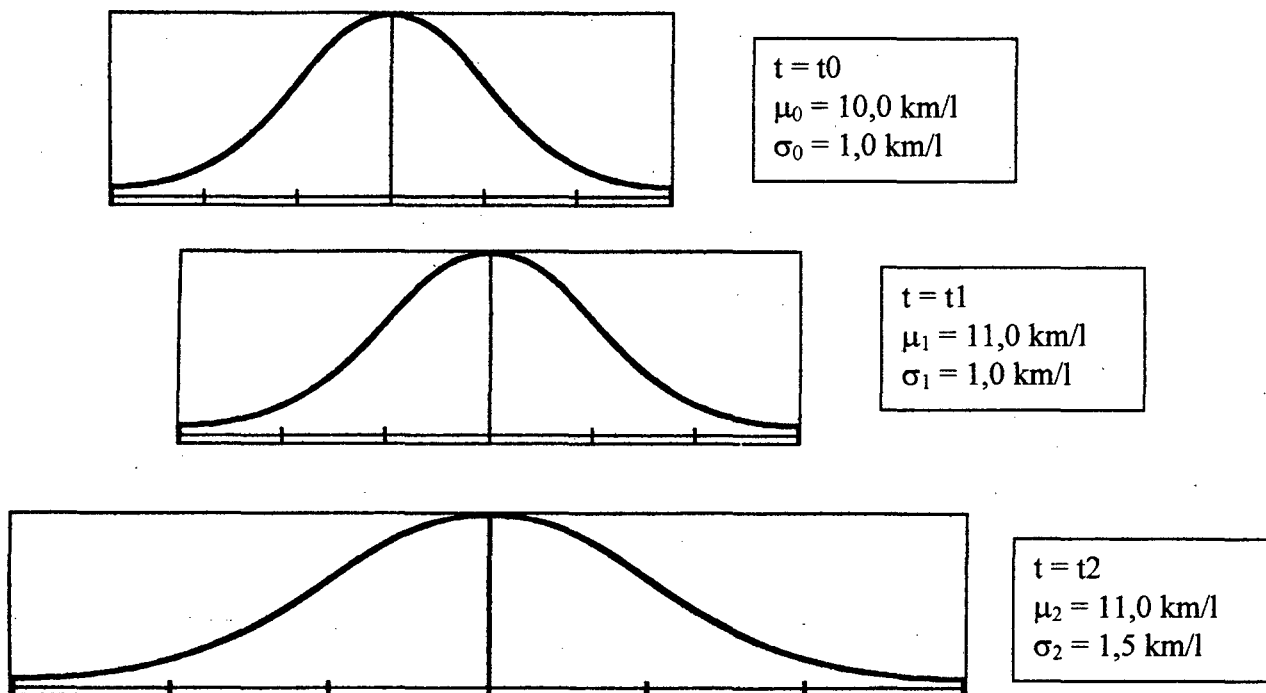


Figura 1-1 – Evolução da função densidade de probabilidade para uma variável de controle considerada, mostrando a variação nos parâmetros da distribuição (média e desvio padrão) devidos a causas especiais de variação.

Damos então a definição de controle estatístico, segundo SHEWHART (1931): “Um fenômeno será dito controlado quando, através da experiência passada, pudermos prever, pelo menos dentro de certos limites, como pode-se esperar que o fenômeno varie no futuro”. Diz-se também que o processo que está sob o efeito de causas comuns apenas está **sob controle estatístico**. A presença de uma causa especial coloca o processo **fora de controle estatístico**.

Como exemplos de causas comuns em processos industriais, temos a interação entre materiais, configuração de máquinas, fadiga do operador, métodos (e outros elementos), sendo bastante difíceis de identificar. Entre exemplos de causas especiais temos máquinas fora de ajuste, diferenças entre fornecedores, métodos alterados, mudança de operador e ambientes de trabalho.

Quando o processo estiver sob controle estatístico (processo estável), podemos avaliar o impacto das alterações introduzidas no mesmo com o objetivo de reduzir a sua variabilidade, que é um dos objetivos do CEP. No nosso exemplo, verificou-se que a regulagem da marcha lenta ajudou a manter a rotação do motor e portanto o desempenho teve uma variabilidade menor.

Por outro lado, a especificação para o desempenho do veículo poderia ser 13,0 km/l e não 10,0 km/l. Dessa forma, notamos que se um processo está sob controle estatístico (agem apenas causas comuns de variação), isto não significa que ele estará atendendo às especificações. Para se obter a relação do CEP com as especificações do produto, é necessário recorrer à **Análise de Capacidade do Processo**, técnica que não abordaremos neste texto. Uma discussão extensiva da capacidade do processo e sua aplicação às indústrias químicas e de processos contínuos é desenvolvida por RAMOS (1999).

1.2.2-) Descrição do Gráfico de Controle

Vamos dar um exemplo da área química (extraído de SCHNEIDER & PRUETT, 1993, pg. 189). Consideremos que dados de 20 carregamentos de negro de fumo são amostrados para se verificar o nível de iodo (I_2) presente. São extraídas amostras de 4 frascos cada e feitas medidas em laboratório.

A figura 1-2 mostra um gráfico (em função do número da amostra) da média do nível de iodo para as 20 amostras tomadas.

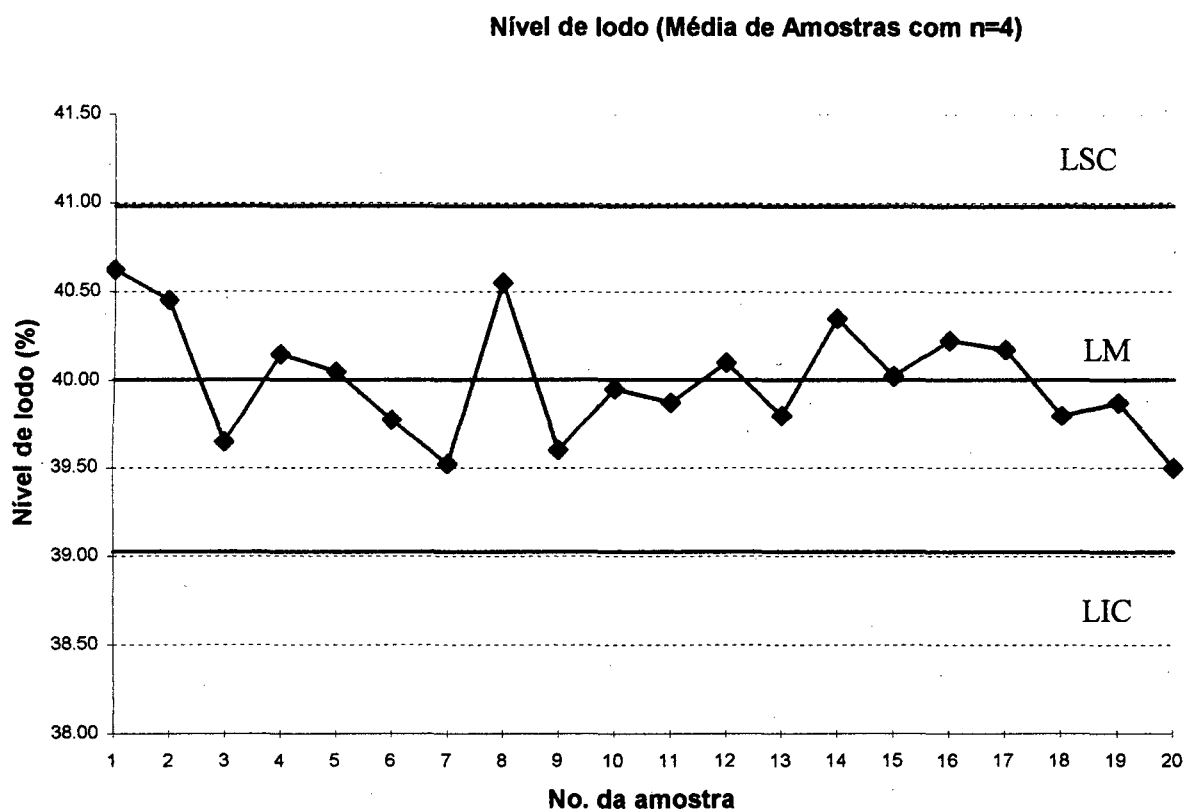


Figura 1-2 – Gráfico de controle para a média da amostra do nível de iodo no negro de fumo.

O gráfico da figura (1-2) mostra valores da variável que representa uma característica da qualidade que desejamos medir (no caso, a média do nível de iodo para as amostras de negro de fumo) para as diferentes amostras. Portanto, em primeiro lugar ele fornece uma idéia da variação dessa variável ao percorrermos as amostras (os valores também poderiam ter sido medidos ao longo do tempo).

O valor médio (ou esperado) para a variável tratada é $X_0 = 40,0$. A **linha média (LM)** representa o valor médio ou esperado da característica da qualidade correspondente ao estado de controle estatístico (quando apenas causas comuns de variação estão presentes). Esta média deverá ser estimada a partir da análise do desempenho do processo durante determinado período de tempo.

O **limite superior de controle (LSC)** e o **limite inferior de controle (LIC)** são colocados geralmente à uma distância de $3\hat{\sigma}$ da linha média, sendo $\hat{\sigma}$ uma estimativa do desvio padrão da

variável representada no gráfico. O uso desses limites, inclusive sua robustez para distribuições não normais, é justificado principalmente pela prática (MONTGOMERY (1996), WHEELER (1994) e DUNCAN (1985)). Podem ser definidos pelo comportamento passado do processo ou calculados a partir dos dados das amostras, não tendo em nenhum caso correspondência com a tolerância da especificação. Mais adiante veremos como são calculados.

De acordo com a seção 1.1, um dos objetivos de um gráfico de controle é o **juízo**, ou seja, testar se um processo está ou não sob controle estatístico. Este conceito tem uma relação com o teste de hipóteses da estatística. Um teste de hipóteses é utilizado quando queremos verificar a validade de uma suposição a respeito de um parâmetro de uma população, utilizando dados de amostras. Definimos a hipótese a ser testada, chamada hipótese H_0 , que será considerada verdadeira a menos que os dados amostrais apresentem evidência significativa em contrário.

Consideremos o teste da média de uma população normal quando a variância é conhecida. A hipótese H_0 é

$$H_0: \mu = \mu_0$$

A hipótese alternativa bilateral fica:

$$H_1: \mu \neq \mu_0$$

E a regra de decisão é:

$$\text{Rejeitar } H_0 \text{ se: } \bar{x} > \mu_0 + Z_{\alpha/2} \sigma_{\bar{x}} \text{ ou } \bar{x} < \mu_0 - Z_{\alpha/2} \sigma_{\bar{x}} \quad (1-1)$$

Onde:

\bar{x} Média amostral

μ_0 Valor de teste

$Z_{\alpha/2}$ Valor da variável normal padrão para o qual $P(Z \leq Z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ Desvio padrão da distribuição de \bar{x} (considerado conhecido)

No caso de controle de processos, é como se fizéssemos a cada amostra um teste da hipótese H_0 de que o processo está sob controle estatístico para a variável considerada. Esta hipótese equivale a dizer que a média da distribuição da variável permanece constante ao longo do tempo, digamos em

um valor μ_0 . Os limites de controle irão equivar aos valores que delimitam a região de aceitação do teste. Portanto, se o valor de uma estatística computada a partir de uma amostra estiver entre os limites de controle, nós não rejeitamos a hipótese de controle estatístico e portanto aceitamos que o processo está sob controle. Se, no entanto, o valor dessa estatística estiver fora dos limites de controle, nós rejeitamos a hipótese de controle estatístico, admitindo que o processo está fora de controle e é válido buscar uma causa especial.

Analogamente ao teste de hipóteses, o gráfico de controle pode falhar, devido à incerteza inerente à estatística. Os erros possíveis são: interpretar processo fora de controle quando na verdade está sob controle (**erro tipo I, com probabilidade α , chamado de alarme falso**) e interpretar processo sob controle quando na verdade está fora de controle (**erro tipo II, com probabilidade β**).

Há diferenças de ponto de vista entre o controle estatístico de processos e os testes de hipóteses, apontadas por MONTGOMERY (1996). O teste de hipóteses trata apenas de desvios no parâmetro testado, não permitindo identificar outras formas de variação como ciclos, tendências ascendentes, estratificação e outros padrões de comportamento não aleatórios que sinalizam que o processo está fora de controle estatístico. Tratamos destes casos a seguir.

1.2.3-) Interpretação do Gráfico e Regras de Sensibilidade

De acordo com a seção 1.2.2, um gráfico de controle indica uma condição de falta de controle quando um ou mais pontos caem fora dos limites superior e inferior. No entanto, se os pontos do gráfico apresentarem padrões de comportamento não aleatórios, mesmo estando todos dentro dos limites, isto também evidenciará uma mudança na distribuição da variável de controle. Entre esses padrões podemos enumerar ciclos, desvios da média, tendências (ascendentes e descendentes) e estratificações (combinação de várias fontes numa mesma amostra). ^{Figura 1-3 (a)} Pode ocorrer também uma mistura desses padrões. Cada tipo de padrão está relacionado a causas especiais específicas, cuja identificação requer experiência e conhecimento do processo. Para o reconhecimento desses padrões não aleatórios utilizam-se geralmente **regras de sensibilidade**.

As regras de sensibilidade são na verdade testes de aleatoriedade. A idéia é que determinados padrões tem uma probabilidade muito baixa de ocorrerem se a premissa de aleatoriedade for satisfeita. Portanto, a sua ocorrência evidencia falta de aleatoriedade e portanto falta de controle nos

dados. O desenvolvimento teórico de testes de aleatoriedade pode ser encontrado em DUNCAN (1985).

Algumas regras de sensibilidade usuais são apresentadas em MONTGOMERY (1996, pág. 149):

1. Um ponto além da distância de 3σ da linha central (“limites de controle”).
2. Dois entre três pontos consecutivos além da distância de 2σ da linha central.
3. Quatro entre cinco pontos consecutivos além da distância de 1σ da linha central.
4. Oito pontos consecutivos do mesmo lado da linha média.

Na verdade, para a regra 4 o ideal seria utilizar a mediana, que no caso da distribuição normal coincide com a média. Há ainda outras regras de sensibilidade utilizadas na prática:

5. Seis pontos consecutivos em uma seqüência ascendente ou descendente.
6. Quinze pontos consecutivos a uma distância menor que 1σ da linha central.
7. Quatorze pontos consecutivos alternadamente ascendentes e descendentes.
8. Um padrão nos dados não usual ou não aleatório.

1.2.4-) Diretrizes para Escolha de Limites de Controle, Tamanho da Amostra e Freqüência de Amostragem

Abordamos nesta seção as diretrizes para a seleção de três parâmetros básicos em um gráfico de controle: limites de controle, tamanho da amostra e freqüência de amostragem.

Limites de Controle

Como vimos na seção 1.2.2, os limites de controle são calculados com base na distribuição normal, utilizando o modelo de testes de hipóteses. A sua escolha é crítica no projeto do gráfico de controle, pois eles dão um balanço entre as probabilidades de ocorrência dos dois erros (tipo I e tipo II) na interpretação do gráfico. Com efeito, ao alargarmos os limites, a probabilidade de que um

ponto caia fora dos limites quando o processo está sob controle diminui (probabilidade do erro tipo I é menor), porém a probabilidade de que um ponto caia entre os limites quando o processo está fora de controle aumenta (probabilidade do erro tipo II é maior).

Os limites 3σ (à distância de três desvios padrão da linha média) são os mais amplamente utilizados na prática. Se for utilizada a distribuição normal, esses limites estão associados à probabilidade $\alpha = 0,27\%$ ($\alpha/2 = 0,135\%$) do erro tipo I, pois $Z_{\alpha/2} = 3$ neste caso. Se uso se deve principalmente à simplicidade de utilizar um múltiplo inteiro do desvio padrão e porque implicam em uma probabilidade α do erro tipo I pequena (menor que 1%) para a maioria das distribuições encontradas na natureza, sendo portanto robustos a desvios de normalidade. (MONTGOMERY (1996), WHEELER (1994) e DUNCAN (1985)).

Outra alternativa é escolher os limites com base no erro tipo I que pode ser cometido (limites probabilísticos). Por exemplo, pode-se adotar o erro tipo I com probabilidade $\alpha/2 = 0.001$, o que fornece um limite de $3,09 \sigma$.

Às vezes são também utilizados **limites de advertência**, que ficam a uma distância de 2σ da linha central. Se um ponto cair fora ou próximo destes limites recomenda-se tomar dados adicionais para verificar se houve desvios no processo. A probabilidade de um ponto cair fora destes limites é de aproximadamente 0,05 (5%).

Ressaltamos ainda que, segundo a fórmula (1-1), os limites dependem do tamanho da amostra e podem variar conforme variarmos o número de elementos de uma amostra, sendo mais estreitos para amostras maiores, pela proporcionalidade com o inverso da raiz quadrada de n .

Tamanho da Amostra

As referências utilizadas neste capítulo demonstram que o gráfico de controle é mais eficiente na detecção de desvios na média do processo quando se utilizam amostras com n maior. Considerando um desvio na média do processo (de k unidades do desvio padrão), quanto maior o tamanho da amostra n , maior será a probabilidade de detectar o desvio na primeira observação tomada após o desvio (a probabilidade β do erro tipo II é menor).

Logo, amostras maiores podem ser tidas por melhores. No entanto, devemos lembrar que a amostra deve ser homogênea (pelo conceito de subgrupos racionais, a ser abordado na próxima

seção), o que é mais difícil a medida que n aumenta; e que o custo aumenta conforme o tamanho da amostra. Na prática, faz-se um compromisso entre o tamanho da amostra e o tamanho do desvio que se quer detectar. Amostras de tamanho 4 a 6 são comumente utilizadas nas indústrias de partes e peças.

Nas indústrias de processos contínuos é bastante comum a escolha de amostras com $n = 1$. As medições são geralmente feitas através de análise de composição química como, por exemplo, as análises de laboratório que são custosas e demoradas. Em certos casos, o procedimento adotado para controle consiste em misturar amostras tomadas em intervalos de tempo sucessivos (para limitar o custo da análise química), considerando a medida da mistura como um valor individual. Além disso, a coleta de amostras é dificultada pela forma de armazenamento dos materiais (tanques e/ou dutos fechados), o que recomenda colher uma amostra com $n = 1$. Como os materiais são frequentemente homogêneos, não é viável coletar uma amostra e dividi-la em partes, pois estaríamos analisando o erro de medição (SCHNEIDER & PRUETT, 1993).

Frequência de amostragem

A frequência de amostragem é um problema complexo, que está associado com diversos fatores como o custo de amostragem, perdas devido à ausência de controle, taxa de produção e probabilidade de ocorrerem desvios no processo. Mais adiante neste texto voltamos a este problema para algumas situações específicas.

Na indústria de processos contínuos, a frequência de amostragem deve ser dimensionada, pela experimentação, em função de certos parâmetros do processo (SCHNEIDER & PRUETT, 1993), que são o *tempo morto* e a sua *constante de tempo*. O tempo morto é o período no qual nenhuma mudança significativa ocorre na medida da performance do processo em decorrência de uma mudança (intencional ou não) na variável do processo. Após o término do tempo morto, a medida da performance do processo começa a se alterar, porém pode decorrer um tempo considerável até que esta mudança seja efetivada. A constante de tempo é o tempo requerido para se efetivar 62,3% (aproximadamente 2/3) da mudança definitiva. Portanto, não faz sentido tomar uma amostra antes do sistema de medição ter chance de demonstrar eventuais alterações, ou seja, deve-se

respeitar esta “inércia do processo”. São encontradas, em processos contínuos envolvendo reações químicas e misturas de componentes, constantes de tempo desde 10 minutos até 4 horas.

1.2.5-) Amostragem Racional e Subgrupos Racionais

Os conceitos de amostragem racional e subgrupos racionais formam a base conceitual dos gráficos de Shewhart. Se os mesmos não forem utilizados, os gráficos obtidos não terão valor.

Para a determinação dos limites de controle do processo é necessária uma estimativa da variabilidade do processo (pois a mesma não é conhecida a priori), sendo esta medida através de uma estatística que represente a variabilidade dentro dos subgrupos. A amostragem racional consiste em selecionar subgrupos tais que **causas especiais se manifestem como uma variação entre subgrupos e não dentro de um subgrupo**, pois se uma causa especial se manifestasse dentro de um subgrupo, a variabilidade estimada com este subgrupo poderia ser muito grande, gerando limites largos que provavelmente não seriam capazes de detectá-la.

Se causas especiais estiverem presentes no processo, a probabilidade de ocorrer **variação entre os subgrupos deverá ser a maior possível** e a probabilidade de ocorrer **variação dentro do subgrupo deverá ser a menor possível**. Para tanto, recomenda-se que a coleta e organização dos dados em subgrupos seja cuidadosa e racional, formando subgrupos homogêneos. Dados de um mesmo operador, mesmo turno, mesma máquina, etc. são critérios eficazes para a seleção de subgrupos racionais. Por exemplo, pode-se amostrar uma seqüência de unidades produzidas no tempo desde que se tome o cuidado de não misturar turnos de produção num mesmo subgrupo.

Entre os problemas relacionados à amostragem enfrentados pelas indústrias de processos contínuos, segundo SCHNEIDER & PRUETT (1993), SHAW (1991) e PYSDEK (1992), temos:

- *a decisão do que medir, como e quando medir*: pela dificuldade de medir determinada variável ou pela necessidade de se ter mais tempo hábil para realizar ajustes no processo, pode-se optar por buscar variáveis alternativas que sejam correlacionadas positivamente com a variável de interesse.
- *o risco de se subestimar a variabilidade do processo*: ao se tomar uma amostra com $n > 1$ de um material homogêneo, os dados dos elementos da amostra serão quase

idênticos entre si a menos de erros de medição. A variabilidade média das amostras será muito pequena neste caso, causando uma subestimação da variabilidade do processo. Os limites de controle ficarão muito estreitos e indicarão causa especial mesmo quando não houver nenhuma. A solução adotada é a mesma que propusemos no tópico anterior sobre tamanho da amostra, ou seja tirar a média das observações de cada amostra interpretando-se como valores individuais.

- *O risco de se superestimar a variabilidade de um processo: pela **estratificação***, que é a tomada de amostras que combinam fontes com variação distinta. Por exemplo, consideremos o caso em que as saídas de duas operações paralelas são combinadas em um ponto comum antes da amostragem e em uma amostra não é possível identificar de onde veio cada unidade do produto. Neste caso as diferenças entre as operações permanecerão ocultas e a variabilidade da amostra tenderá a ser maior em função exatamente dessas diferenças. O efeito é uma superestimação da variabilidade do processo, levando a limites largos com os quais não se é capaz de detectar causas especiais. Para resolver o problema é necessário controlar cada processo com um gráfico separado. Uma situação análoga que leva ao mesmo problema é a da **mistura de processos**, que consiste na tomada de amostras sucessivas a partir de fontes com variação distinta.

1.3-) Gráfico de Controle para Variáveis

Os gráficos de controle são classificados em função da característica da qualidade que queremos controlar. Se esta característica for resultado de uma medição (por exemplo: diâmetro de uma peça), foi mostrado por SHEWHART (1931) que na prática ela pode ser bem controlada por uma medida de tendência central e uma de variabilidade (não sendo necessários outros parâmetros como “skewness” e “kurtosis”), independentemente da distribuição da característica da qualidade em estudo. Para analisar estas características utilizam-se **gráficos de controle para variáveis**. Se, ainda, a característica da qualidade for resultado de uma classificação ou contagem (por exemplo: porcentagem de unidades defeituosas) utiliza-se o **gráfico de controle para atributos**. Neste

trabalho iremos nos concentrar apenas nos gráficos de controle para variáveis, que são aqueles utilizados dentro do problema que desejamos estudar.

Como especificado acima, no gráfico de controle para variáveis é necessário controlar o valor médio e a variabilidade. Para controlar o valor médio usamos a média amostral \bar{X} . Para o controle da variabilidade pode-se utilizar o gráfico de R (amplitude), S (desvio padrão) ou RM (amplitude móvel).

O uso da estatística R é histórico no controle de processos. Além de ser mais fácil de calcular do que a estatística S (considere-se o tempo em que não havia calculadora), é um estimador tão eficiente quanto S para a variabilidade do processo para pequenas amostras (DUNCAN, 1985). A eficiência relativa de dois estimadores θ_1 e θ_2 não viesados de um parâmetro θ é dada por (NEWBOLD, 1995):

$$\text{Eficiência Relativa} = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} \quad (1-2)$$

Onde $\hat{\theta}_1$ e $\hat{\theta}_2$ são estimadores não viesados de θ .

Para amostras de tamanho 10 ou maiores, a estatística R perde eficiência em relação a S como estimador da variabilidade do processo (MONTGOMERY, 1996, pg. 212) e portanto é mais recomendável utilizar o gráfico de S. Isto é intuitivo, pois para amostras grandes o uso da amplitude nos leva a perder informações sobre a amostra ao considerar apenas o valor das observações máxima e mínima da amostra no seu cálculo.

Quando $n = 1$, uma forma de se calcular a amplitude é comparar a observação atual com a anterior, obtendo-se uma **amplitude móvel (RM)**.

Os diferentes tipos de gráficos de controle para variáveis são resumidos na Tabela 1-1 e serão explicados com maior detalhe nos itens a seguir.

Tamanho da amostra	Tipo de gráfico
$n \geq 10$	$\bar{X}; S$
$2 \leq n \leq 9$	$\bar{X}; R$
$n = 1$	$X; RM$

Tabela 1-1 - Tipos de gráfico de controle para variáveis em função do tamanho das amostras.

Como no problema que desejamos estudar (processos contínuos) as amostras são geralmente pequenas ($n < 10$), senão aquelas com $n = 1$, não iremos abordar os gráficos \bar{X} e S neste texto.

A seguir são dados os métodos para se construir os gráficos de controle \bar{X} e R e X e RM, além da interpretação dos gráficos e análise de sua performance.

1.3.1-) Gráficos de Controle para Média e Amplitude

Os gráficos de controle para a média (\bar{X}) e amplitude (R) são gráficos de controle para variáveis utilizados quando os subgrupos de amostragem têm tamanho maior do que 1 e menor do que 10. O desenvolvimento completo das fórmulas para estes gráficos pode ser encontrado em DUNCAN (1985), entre outras referências. Damos aqui as principais fórmulas utilizadas.

Os valores da média (μ) e do desvio padrão (σ) da característica da qualidade em estudo geralmente não são conhecidos de antemão. Os gráficos de \bar{X} e R utilizam então as médias das amostras (\bar{X}) para estimar a média do processo (μ) e as amplitudes das amostras (R) para estimar o

desvio padrão (σ) do mesmo.

Ao iniciar o controle do processo os parâmetros podem ser estimados através das estatísticas obtidas entre as primeiras amostras coletadas. O desvio padrão σ é estimado utilizando-se a amplitude média de todas as amostras, sendo d_2 é uma constante dependente de n (ver DUNCAN 1985):

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (1-3)$$

Para a construção dos gráficos de controle, consideremos m amostras de tamanho n , obtidas de acordo com o conceito de subgrupos racionais. Obtém-se inicialmente a média de cada amostra pela fórmula (1-4) abaixo:

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \dots + x_{in}}{n}, \quad i = 1, 2, \dots, m \quad (1-4)$$

Deve-se calcular também a amplitude de cada amostra, que é a diferença absoluta entre o valor máximo e o valor mínimo:

$$R_i = |x_{i,\max} - x_{i,\min}|, \quad i = 1, 2, \dots, m \quad (1-5)$$

Feito isto, obtém-se a média das médias e a média das amplitudes, que são dados em (1-6) e (1-7):

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (1-6)$$

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (1-7)$$

Por fim, calculam-se os limites de controle – **Limite Superior de Controle (LSC)** e **Limite Inferior de Controle (LIC)** – e a **Linha Média (LM)** para ambos os gráficos. Para o gráfico de \bar{X} temos os seguintes limites de 3σ (observar que para a distribuição de \bar{X} o desvio padrão é igual ao desvio padrão da distribuição de X , dividido pela raiz quadrada de n):

$$\begin{aligned} \text{LSC} &= \bar{\bar{x}} + \frac{3\hat{\sigma}}{\sqrt{n}} = \bar{\bar{x}} + A_2 \bar{R} \\ \text{LM} &= \bar{\bar{x}} \end{aligned} \quad (1-8)$$

E para o gráfico de R:

$$\begin{aligned} \text{LSC} &= \bar{R} + 3 \hat{\sigma}(R) = D_4 \bar{R} \\ \text{LM} &= \bar{R} \end{aligned} \quad (1-9)$$

$$\text{LIC} = \bar{R} - 3 \hat{\sigma}(R) = D_3 \bar{R}$$

As constantes A_2 , D_3 e D_4 são tabuladas para valores de n por vários autores, entre eles DUNCAN (1985).

Como vimos na seção 1.2.4, apesar de assumirmos a premissa de normalidade nos dados para o cálculo dos limites de controle, estes limites são bastante robustos com relação à normalidade dos dados, especialmente para amostras de tamanho mínimo 4 ou 5 (em consequência do teorema do limite central, que diz que independentemente da distribuição de uma variável X , a distribuição de \bar{X} tenderá à normal conforme n aumenta).

Uma consideração bastante importante: se após a construção do gráfico R este indicar que o processo está fora de controle, deve-se eliminar as causas especiais evidenciadas por pontos fora dos limites ou padrões não aleatórios e recalcular o \bar{R} (sem chegar a analisar o gráfico de \bar{X}). Este valor de \bar{R} é utilizado para se determinar novos limites de controle e linha média no gráfico R e também novos limites no gráfico de \bar{X} .

Os primeiros limites de controle obtidos neste processo são chamados *limites preliminares*, pois representam um entendimento inicial do processo. Estes limites devem ser revistos após serem detectadas e eliminadas causas especiais, e também devem sofrer revisões periódicas.

Análise da performance dos gráficos de média e amplitude

Nesta seção vamos introduzir um conceito que nos ajudará a fazer a análise da performance dos gráficos \bar{X} e R.

Uma medida de performance dos gráficos de controle é a **ARL** (“**Average Run Length**” – **Comprimento Médio da Seqüência**), que consiste no número médio de amostras a serem observadas até se detectar um ponto fora dos limites de controle. Portanto, a ARL supõe apenas a 1ª regra de detecção de causas especiais, que é a presença de um ponto fora dos limites de controle.

Para um gráfico sob controle estatístico ARL_0 (dita *ARL sob controle*) expressa o número de amostras analisadas até a primeira ocorrência de um alarme falso. Intuitivamente observamos que a ARL deve ser inversamente proporcional à probabilidade de ocorrência do alarme falso, pois quanto maior a probabilidade para este tipo de alarme, espera-se um número menor de observações até que ele aconteça. Com efeito, sua fórmula, conforme [MONTGOMERY (1996), Pg. 209], sendo α a probabilidade de erro tipo I (detectar processo fora de controle quando o mesmo está sob controle), é

$$ARL_0 = \alpha \sum_{i=1}^{\infty} i(1-\alpha)^{i-1} = \frac{1}{\alpha} \quad (1-10)$$

dada por:

Os gráficos de controle baseados em limites de 3-sigma apresentam uma ARL sob controle bastante elevada, em torno de 370, quando a distribuição é normal (pois $\alpha = 0,027$ neste caso), o que significa uma baixa taxa de alarmes falsos, desde que não levemos em conta as demais regras de sensibilidade, que aumentam a taxa de alarmes falsos.

Para o caso de processo fora de controle, ARL_1 (dita *ARL fora de controle*) serve para determinar o número médio de amostras a analisar desde a ocorrência de uma causa especial até a detecção da mesma. Sendo β a probabilidade de erro tipo II (detectar processo sob controle quando o mesmo está fora de controle), a fórmula para a ARL fica [MONTGOMERY (1996), Pg. 209]:

$$ARL_1 = \frac{1}{1-\beta} \quad (1-11)$$

O conceito para ARL sob controle e fora de controle é o mesmo, porém os objetivos são diferentes. No caso de controle estatístico, deseja-se que a ARL_0 seja a maior possível, para que os alarmes falsos sejam bastante infreqüentes. Porém, na situação de falta de controle, a intenção é detectar o problema o mais rapidamente possível e então é interessante que a ARL_1 seja a menor possível.

Para verificarmos a performance do gráfico de controle na detecção de desvios da média do processo em função de σ e do tamanho da amostra, consideremos a expressão teórica (baseada na

distribuição normal) para a probabilidade β do erro tipo II, dada por MONTGOMERY (1996, pág. 206):

$$\beta = \Phi(L - k\sqrt{n}) - \Phi(-L - k\sqrt{n}) \quad (1-12)$$

onde: $\Phi(Z_0) = P(Z < Z_0)$ na distribuição normal padrão

L é o múltiplo de desvios padrões na determinação dos limites de controle (geralmente $L = 3$)

k é a magnitude do desvio ($\mu - \mu_0 = k\sigma$)

n é o tamanho da amostra

O gráfico de \bar{X} , para os tamanhos de amostra de quatro, cinco ou seis, é eficaz para detectar grandes desvios do tipo “shift” sustentado (ou degrau) na média do processo, da ordem de 2σ ou maiores. Para pequenos desvios (menores que $1,5\sigma$), há ainda uma boa chance de que os mesmos sejam detectados rapidamente, embora geralmente não na primeira amostra analisada após o desvio ter ocorrido. Por exemplo, para $n = 5$ e desvio de $1,5\sigma$ ($k = 1,5$), a probabilidade de deixar de detectar tal desvio na primeira amostra, dada pela fórmula (1-12), é:

$$\beta = \Phi(3 - 1,5\sqrt{5}) - \Phi(-3 - 1,5\sqrt{5}) = \Phi(-0,354) - \Phi(-6,354) = 0,361$$

Calculando a ARL_1 nesse caso temos:

$$ARL_1 = \frac{1}{1 - \beta} = \frac{1}{0,639} \cong 1,56$$

Para o caso do gráfico de R, MONTGOMERY (1996) observa que para pequenas amostras ($n = 4, 5$ ou 6) o mesmo é bem pouco eficaz na detecção de desvios na variabilidade do processo com uma razão menor ou igual a 2 entre a variabilidade nova e a antiga do processo. O gráfico R

serve então para detectar desvios realmente grandes, com uma razão da ordem de 3 ou superior entre a variabilidade nova e a variabilidade antiga do processo.

Exemplo de aplicação

Consideremos o exemplo da seção 1.2.2, em que analisamos carregamentos de negro de fumo. Os dados utilizados no exemplo são dados na tabela 1-2.

Amostra	Valores do nível de iodo				\bar{x}	R
	x1	x2	x3	x4		
1	40.9	40.5	40.1	41.0	40.63	0.90
2	40.8	40.3	39.2	41.5	40.45	2.30
3	39.6	39.4	40.3	39.3	39.65	1.00
4	40.6	40.2	40.6	39.2	40.15	1.40
5	40.4	40.4	39.4	40.0	40.05	1.00
6	39.2	40.4	39.8	39.7	39.78	1.20
7	39.9	39.0	39.5	39.7	39.53	0.90
8	40.2	40.2	41.4	40.4	40.55	1.20
9	39.9	39.9	39.2	39.4	39.60	0.70
10	40.5	40.1	40.6	38.6	39.95	2.00
11	39.9	39.2	40.2	40.2	39.88	1.00
12	41.1	38.8	39.9	40.6	40.10	2.30
13	40.1	40.3	39.1	39.7	39.80	1.20
14	39.6	41.0	40.6	40.2	40.35	1.40
15	40.0	38.6	41.5	40.0	40.03	2.90
16	40.5	40.0	40.2	40.2	40.23	0.50
17	40.7	40.2	39.6	40.2	40.18	1.10
18	38.8	40.2	40.3	39.9	39.80	1.50
19	40.6	39.2	40.4	39.3	39.88	1.40
20	39.7	39.0	39.9	39.4	39.50	0.90

Tabela 1-2 – Valores coletados para o nível de iodo num carregamento de negro de fumo com $n = 4$; valores de \bar{X} e R já calculados (fonte SCHNEIDER & PRUETT, 1993, pág. 189).

Os valores médios de \bar{X} e \bar{R} para as 20 amostras são:

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 40,00$$

$$\bar{R} = \frac{\sum_{i=1}^{20} R_i}{20} = 1,340$$

Para $n = 4$, temos os seguintes valores das constantes A_2 , D_3 e D_4 :

$$A_2 = 0,729$$

$$D_3 = 0$$

$$D_4 = 2,282$$

Portanto, os limites de controle ficam:

- Gráfico de \bar{X}

$$LSC = \bar{x} + A_2 \bar{R} = 40,00 + (0,729)(1,340) = 40,98$$

$$LC = \bar{x} = 40,00$$

$$LIC = \bar{x} - A_2 \bar{R} = 40,00 - (0,729)(1,340) = 39,03$$

- Gráfico de R

$$LSC = D_4 \bar{R} = 2,282(1,340) = 3,058$$

$$LC = \bar{R} = 1,340$$

$$LIC = D_3 \bar{R} = 0(1,340) = 0$$

Os gráficos de \bar{X} e R são mostrados a seguir, na Figura 1-3. O gráfico de R deve ser analisado primeiro. Não há nenhum ponto fora dos limites, porém há um ponto bastante próximo ao limite superior de controle, o que nos levaria a fazer uma verificação no processo. Supondo que nada de anormal foi encontrado, podemos analisar o gráfico de \bar{X} , que não apresenta pontos fora dos limites de controle nem padrões de comportamento não aleatórios, indicando que o processo está sob controle.

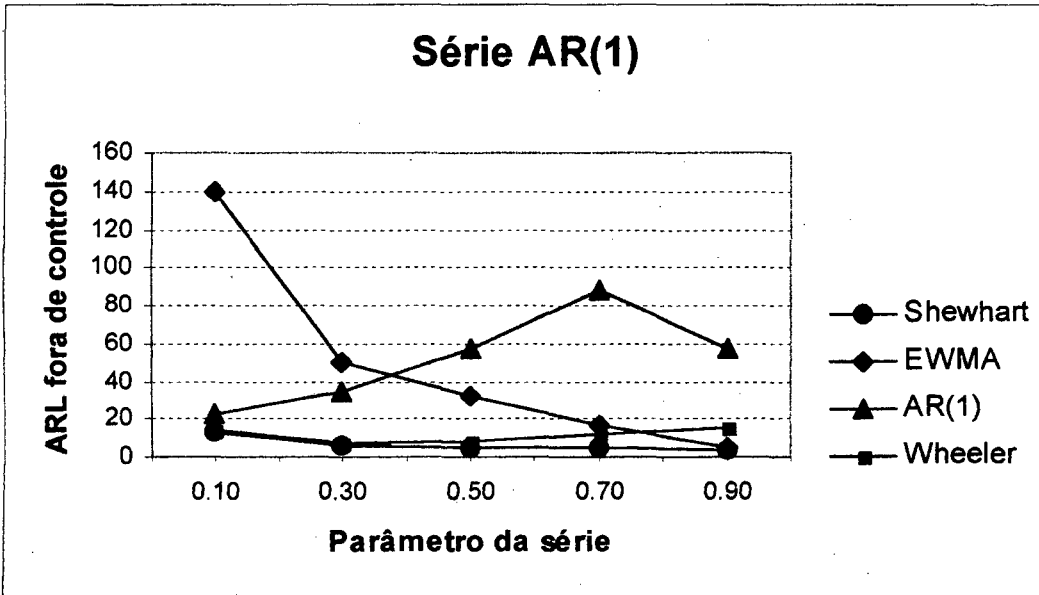
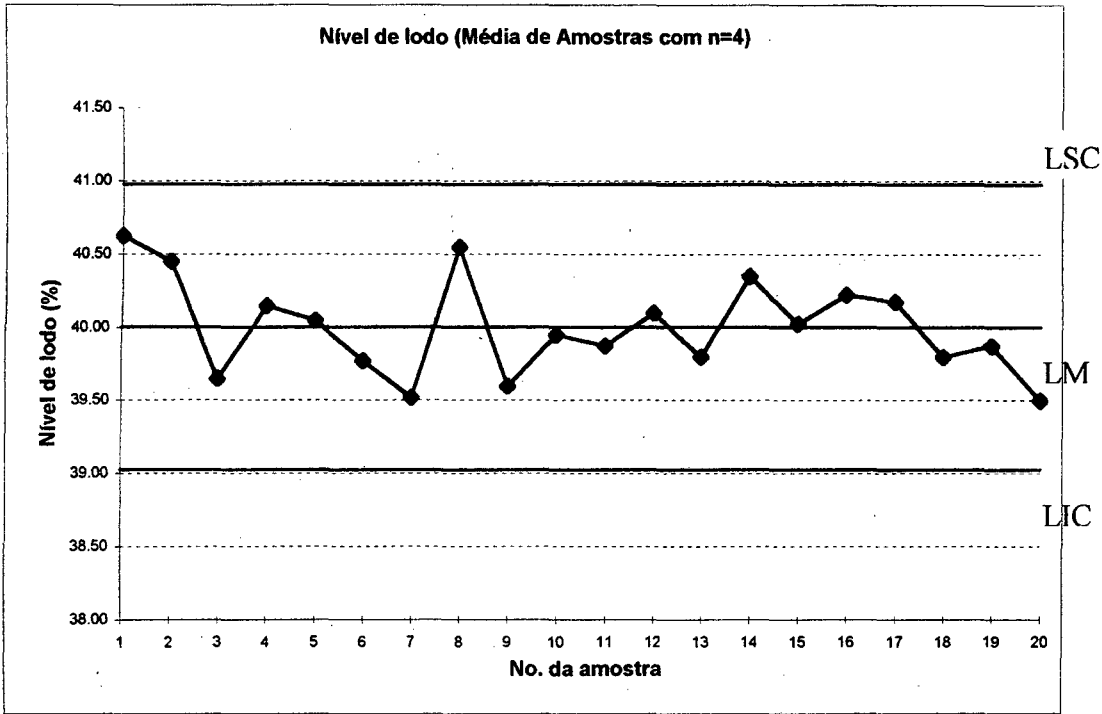


Figura 1-3 – Gráficos de \bar{X} e R para o nível de iodo do negro de fumo.

1.3.2-) Gráficos de Controle para Valores Individuais e Amplitude Móvel

Nas situações em que se utiliza um tamanho de amostra $n = 1$, utiliza-se o gráfico de controle para valores individuais. Como não há possibilidade de se medir a amplitude (R) da amostra com $n = 1$, pode-se utilizar a amplitude móvel de duas observações sucessivas para se estimar a variabilidade do processo. A amplitude móvel (RM) é dada por:

$$RM_i = |x_i - x_{i-1}| \quad (1-13)$$

CRYER & RYAN (1990), sugerem o uso da estatística S calculada a partir de todos os dados e a estimativa do desvio padrão pela razão S/c_4 , sendo c_4 uma constante dependente de n . Este estimador é mais eficiente que RM/d_2 tanto para dados independentes e normais quanto para dados não independentes. Sua recomendação é o cálculo de ambos os estimadores. Se as estimativas forem concordantes, podemos confiar na performance do gráfico de controle; porém, se as estimativas não concordarem, concluiremos que há um problema. O método do RM/d_2 (amplitude móvel), que é o único a adotarmos aqui, é o recomendado em todas as demais fontes de nossa bibliografia básica.

Obtendo-se a amplitude móvel média para o conjunto de observações (exceto a primeira), é possível calcular os limites de controle para o gráfico de valores individuais de X:

$$\begin{aligned} LSC &= \bar{x} + \frac{3 \overline{RM}}{d_2} \\ LM &= \bar{x} \\ LIC &= \bar{x} - \frac{3 \overline{RM}}{d_2} \end{aligned} \quad (1-14)$$

Sendo $d_2 = 1,128$ (atribuído para $n = 2$).

É comum utilizar também o gráfico de controle para a amplitude móvel. Os limites são os mesmos que para o gráfico de R (considerando-se $n = 2$) e suas fórmulas são reproduzidas para maior comodidade:

$$\begin{aligned}
 LSC &= D_4 \overline{RM} \\
 LM &= \overline{RM} \\
 LIC &= D_3 \overline{RM}
 \end{aligned}
 \tag{1-15}$$

No próximo parágrafo tratamos da interpretação dos gráficos X e RM, performance e cálculo da ARL e considerações sobre a normalidade dos dados.

Interpretação dos gráficos

O gráfico para Valores Individuais é interpretado da mesma forma que um gráfico de \bar{X} (supondo-se a independência das amostras).

Deve-se ter cuidado ao utilizar regras de sensibilidade para a interpretação do gráfico de RM, porque as amplitudes móveis são autocorrelacionadas (não independentes) e isto pode frequentemente induzir tendências ou ciclos no gráfico.

Performance e cálculo da ARL

A performance do gráfico de controle para valores individuais na detecção de pequenos e médios desvios na média do processo é bastante ruim, com a ARL chegando a 44 amostras ($\beta = 0,9772$) para a detecção de um desvio de 1σ na média. Isto pode ser verificado substituindo em nossa fórmula anterior (1-12) os valores $L=3$, $k=1$ e $n=1$ para obter o valor de β e calculando a ARL conforme (1-11). Se a frequência de amostragem for 1 hora, serão decorridos quase dois dias até o gráfico sinalizar desvios dessa ordem. Para aumentar a performance do gráfico na detecção de pequenos desvios na média sugere-se utilizar outros gráficos, como CUSUM e EWMA (MONTGOMERY, 1996). Para desvios maiores (da ordem de 3-sigma) o gráfico tem uma melhor performance, sendo a $ARL = 2$ ($\beta = 0,5$) para $k=3$.

Normalidade dos dados

As considerações que fizemos acima supõem que há normalidade dos dados e os resultados podem ser bastante afetados no caso de $n = 1$ se os dados estiverem afastados da distribuição normal, pois o teorema do limite central não é aplicável como no caso de \bar{X} . Nesse caso, uma solução é a transformação da característica não normal em outra que seja aproximadamente normal e então construir os gráficos de controle. Outra possibilidade é o uso de limites probabilísticos como os de probabilidade $\alpha/2 = 0,001$ (0,1%).

Exemplo de aplicação 1.2

Os dados referem-se a valores da compressão de pelotas queimadas (de $-16,0$ a $+12,5$ mm) e foram obtidos a cada duas horas.

<u>Observação</u>	<u>Comp. Pelotas</u>	<u>R móvel</u>
1	282	-----
2	288	6
3	260	28
4	263	3
5	297	34
6	256	41
7	279	23
8	279	0
9	255	24
10	282	27
11	255	27
12	278	23
13	270	8
14	258	12
15	243	15
16	249	6
17	294	45
18	301	7
19	280	21
20	290	10
21	286	4
22	333	47
23	330	3
24	298	32
25	335	37
26	306	29
27	300	6
28	306	6
29	293	13
30	302	9

Tabela 1-3 – Valores coletados para a compressão de pelotas, com amostras de $n = 1$; valores de RM já calculados. (fonte: “CEP Para Processos Contínuos e em Bateladas – Estudos de Caso”, Apostila de treinamento do Prof. Alberto Ramos)

Os valores médios de X e RM para as 30 amostras (29 no caso de RM) são:

$$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{30} = 284,9$$

$$\overline{RM} = \frac{\sum_{i=1}^{14} RM_i}{29} = 18,8$$

Para $n = 2$ (amplitudes móveis de ordem 2), temos os seguintes valores das constantes d_2 , D_3 e D_4 :

$$d_2 = 1,128$$

$$D_3 = 0$$

$$D_4 = 3,267$$

Portanto, os limites de controle dos dois gráficos ficam:

- Gráfico de X

$$LSC = \bar{x} + \frac{3 \overline{RM}}{d_2} = 284,9 + \frac{(3)(18,8)}{(1,128)} = 335,0$$

$$LC = \bar{x} = 284,9$$

$$LIC = \bar{x} - \frac{3 \overline{RM}}{d_2} = 284,9 - \frac{(3)(18,8)}{(1,128)} = 234,9$$

- Gráfico de RM

$$LSC = D_4 \overline{RM} = 3,267(18,8) = 61,5$$

$$LC = \overline{RM} = 18,8$$

$$LIC = D_3 \overline{RM} = 0(18,8) = 0$$

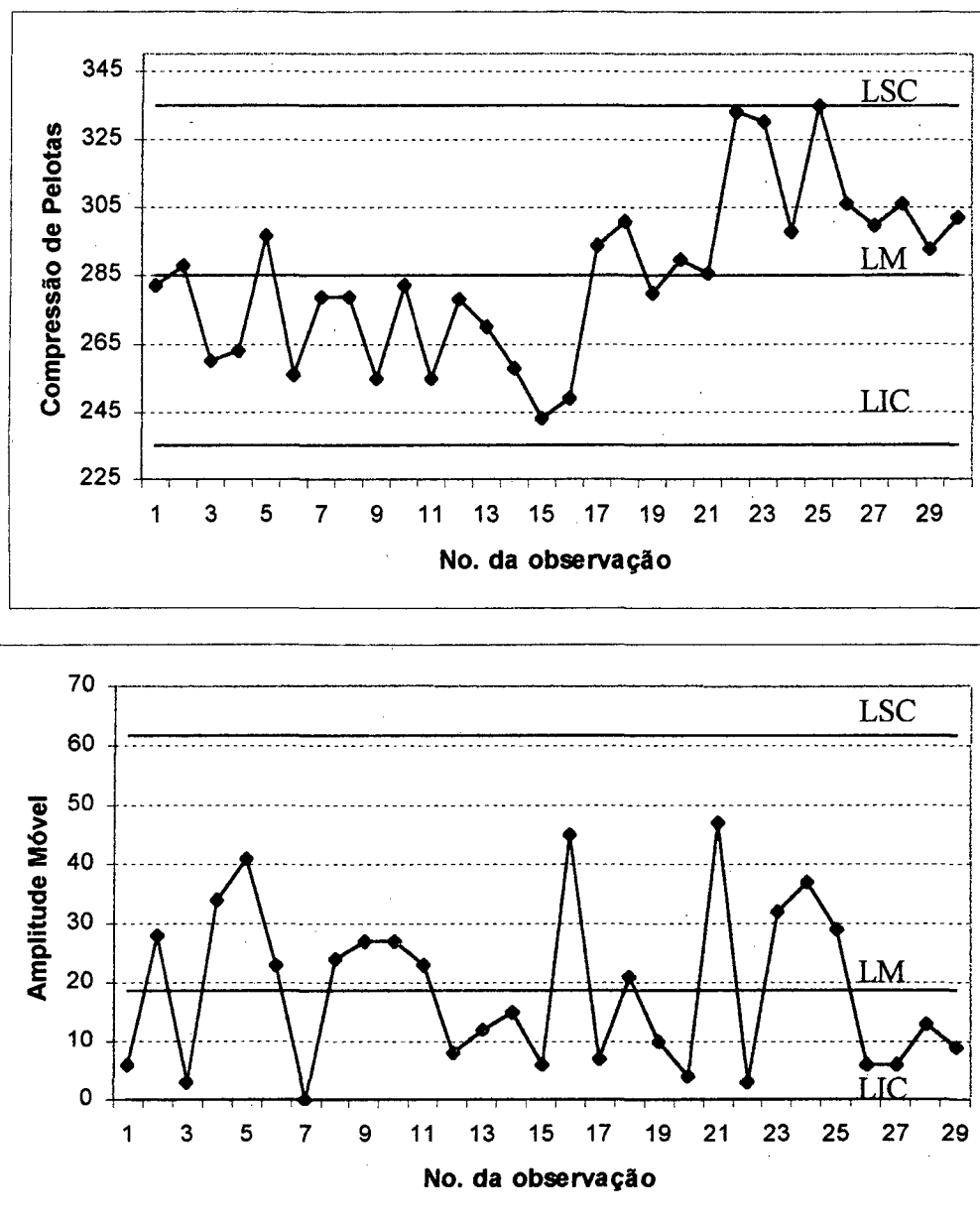


Figura 1-4 – Gráficos de X (valores individuais) e RM para a compressão de pelotas.

Os gráficos de X e RM são mostrados na Figura 1-4. Observamos que não há nenhum ponto fora dos limites de controle nem padrões não aleatórios no gráfico de RM. Passando a analisar o gráfico de X, o mesmo indica claramente uma condição de ausência de controle estatístico, correspondente a um desvio na média do processo. Como podemos observar, pontos consecutivos abaixo da linha média são sucedidos por pontos consecutivos acima da linha média após a

observação 16, o que indica alteração na média do processo, que é uma causa especial a ser buscada e solucionada.

Capítulo 2 – Autocorrelação

2.1-) Correlação Linear Entre Duas Variáveis

Em nosso trabalho estamos tratando do controle estatístico aplicado a processos contínuos. Uma característica bastante comum dos processos contínuos é a **falta de independência** ou a **presença de autocorrelação nos dados observados**. Neste capítulo iremos conceituar a autocorrelação e mostrar alguns exemplos da mesma.

O conceito de autocorrelação é derivado do conceito de **correlação**. A correlação trata da relação linear existente entre duas variáveis aleatórias. O índice que mede o grau da associação linear entre duas variáveis aleatórias é denominado **coeficiente de correlação linear**, dado pela fórmula abaixo:

$$\rho = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad (2-1)$$

Demonstra-se que o coeficiente de correlação deverá estar entre -1 e 1 , ou seja:

$$-1 \leq \rho \leq 1 \quad (2-2)$$

Para o coeficiente de correlação ρ , são válidas as seguintes interpretações:

- (i) Uma correlação de -1 implica uma associação linear negativa perfeita.
- (ii) Uma correlação de 1 implica uma associação linear positiva perfeita.
- (iii) Uma correlação de 0 implica que não há associação linear entre as variáveis.
- (iv) Quanto maior o valor absoluto da correlação, maior será a associação linear entre as variáveis aleatórias.

Quando o coeficiente de correlação entre duas variáveis não é conhecido, pode-se tomar uma amostra de n valores das variáveis x e y e calcular-se o **coeficiente de correlação da amostra (r)**:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-3)$$

Como exemplo, fornecemos na tabela 2-1 pesos e alturas de 33 mulheres, extraído de WHEELER (1994).

Peso (Kg)	Altura (m)	Peso (Kg)	Altura (m)	Peso (Kg)	Altura (m)
42,2	1,52	45,4	1,55	48,6	1,47
47,7	1,52	46,8	1,60	50,8	1,57
46,8	1,57	47,2	1,63	56,8	1,50
55,4	1,55	53,6	1,60	55,4	1,57
56,3	1,63	56,8	1,65	53,6	1,63
55,4	1,68	61,7	1,52	59,0	1,60
61,3	1,57	60,8	1,65	57,7	1,63
62,2	1,63	61,3	1,68	59,9	1,68
63,6	1,55	66,3	1,57	65,4	1,60
65,8	1,63	67,2	1,70	66,3	1,70
70,8	1,60	70,8	1,63	69,0	1,68

Tabela 2-1 – Pesos (em Kg) e Alturas (em metros) de 33 mulheres.

No gráfico de dispersão (figura 2-1) com os pesos no eixo das abscissas e as alturas no eixo das ordenadas, podemos observar que o peso e a altura das mulheres apresentam uma correlação positiva, pois em geral para pesos maiores correspondem alturas maiores e vice-versa.

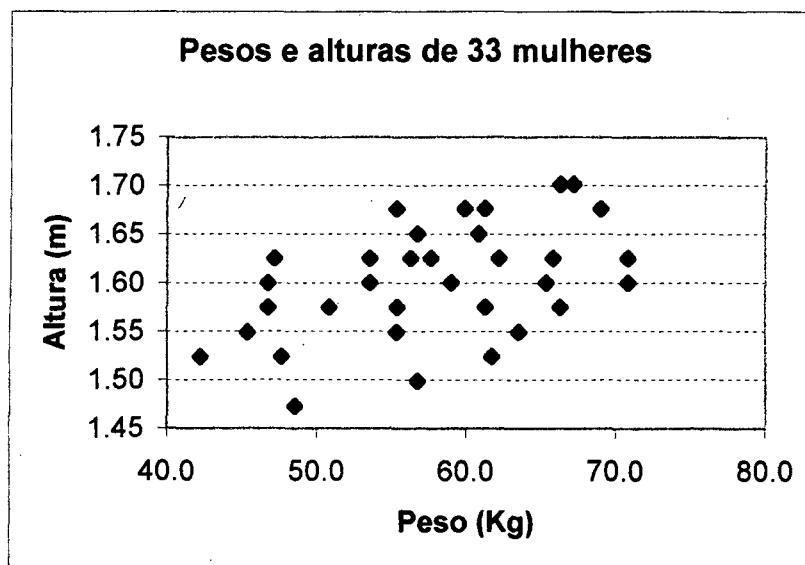


Figura 2-1 – Gráfico de dispersão para os pesos e alturas de 33 mulheres.

Com efeito, o coeficiente de correlação entre as duas variáveis, calculado segundo a fórmula (2.3), vale $r = 0,4689$.

2.2-) Autocorrelação

Nos processos contínuos há geralmente uma relação entre valores sucessivos de uma seqüência de observações no tempo. Isto se deve ao fato de que os materiais contínuos, sendo geralmente homogêneos, apresentam uma certa “inércia” para mudar de estado, conforme visto na seção 1.2.4. Ou seja, as mudanças observáveis no processo não são repentinas, mas ocorrem gradualmente. Desta forma, há uma grande chance de que a observação do processo em um dado instante esteja próxima das observações em instantes anteriores.

Analogamente ao coeficiente de correlação, que é usado para caracterizar o relacionamento entre duas variáveis, o coeficiente de autocorrelação é usado para caracterizar o relacionamento entre observações da mesma variável, ou seja, se as observações consecutivas são ou não correlacionadas.

Numa seqüência de valores da variável x , o coeficiente de autocorrelação de “lag k ” ou “retardo k ” é obtido definindo-se uma segunda variável com as observações da variável x defasadas

de k unidades de tempo. Na tabela 2-2 é dada uma seqüência deste tipo, com as seqüências correspondentes com lag $k=1$ e $k=2$ colocadas ao lado. Os dados da tabela 2-2 correspondem a medidas de peso molecular de um polímero, feitas a cada duas horas (extraído de MONTGOMERY (1996), pg. 426).

t	x(t)	x(t-1)	x(t-2)
1	20.48	-	-
2	20.25	20.48	-
3	20.17	20.25	20.48
4	19.95	20.17	20.25
5	19.83	19.95	20.17
6	19.43	19.83	19.95
7	19.40	19.43	19.83
8	19.47	19.40	19.43
9	19.72	19.47	19.40
10	19.83	19.72	19.47
11	19.35	19.83	19.72
12	19.48	19.35	19.83
13	19.66	19.48	19.35
14	19.54	19.66	19.48
15	19.70	19.54	19.66
16	20.39	19.70	19.54
17	20.15	20.39	19.70
18	20.21	20.15	20.39
19	20.10	20.21	20.15
20	20.12	20.10	20.21

Tabela 2-2 – Seqüência de valores de peso molecular de um polímero, acompanhada de séries com os mesmos valores defasados com lag $k=1$ e $k=2$.

O cálculo do coeficiente de autocorrelação é feito através da fórmula utilizada no caso da correlação linear, com a diferença de que agora as variáveis são x_t e x_{t-k} ao invés de x e y .

$$r_k = \frac{\sum_{t=k+1}^n (x_t - M_1)(x_{t-k} - M_2)}{\sqrt{\sum_{t=k+1}^n (x_t - M_1)^2 \sum_{t=1}^{n-k} (x_t - M_2)^2}} \cong \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (2-4)$$

A fórmula acima envolve as médias M_1 e M_2 , que não são exatamente as mesmas, pois elas são baseadas em $(n-k)$ pontos que não são exatamente os mesmos devido à defasagem entre as seqüências. No denominador da fórmula aparecem dois desvios padrão que também não são os mesmos pelas mesmas razões. Como a seqüência com a qual estamos trabalhando é uma só, é usual simplificar estas fórmulas fazendo as seguintes suposições, que levam à fórmula aproximada da direita:

1. As duas médias M_1 e M_2 podem ser estimadas utilizando-se a média convencional de todos os n valores.
2. Os dois desvios padrão do denominador podem ser estimados pelo desvio padrão de todos os n valores.

Os valores da autocorrelação para o exemplo da tabela 2-2, calculados segundo a fórmula (2.4), para os lags $k=1$ e $k=2$, são:

$$r_1 = 0,649$$

$$r_2 = 0,379$$

Os coeficientes de autocorrelação podem ser observados graficamente, construindo-se um gráfico de x_t (também grafado como $x(t)$) em função de x_{t-k} (também grafado como $x(t-k)$). Nas Figuras 2-2 e 2-3 são apresentados os gráficos para $k=1$ e $k=2$.

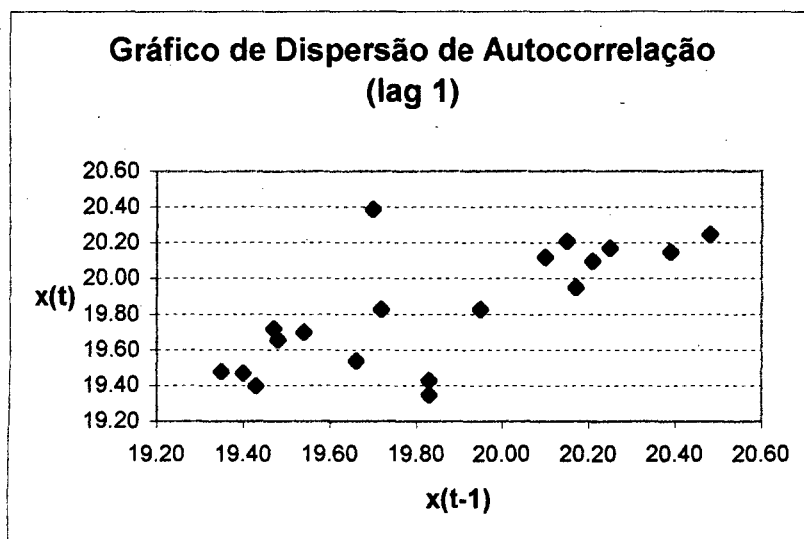


Figura 2-2 – Gráfico de dispersão de autocorrelação (lag 1) para o peso molecular de um polímero.

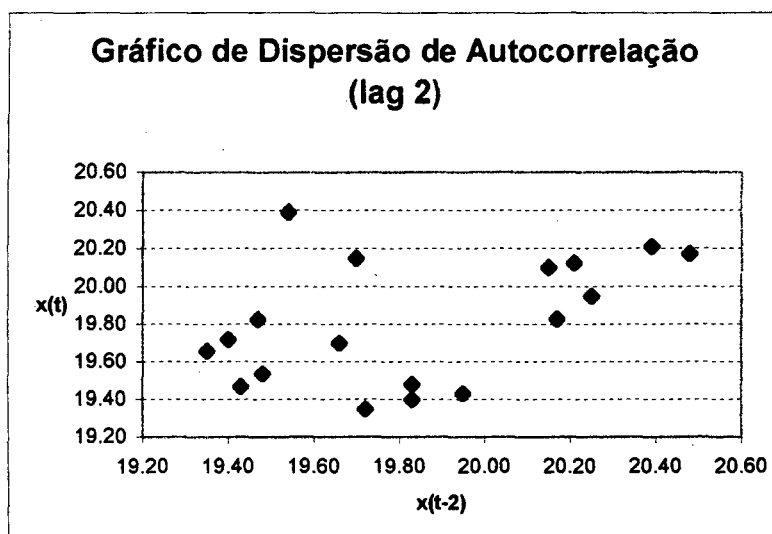


Figura 2-3 – Gráfico de dispersão de autocorrelação (lag 2) para o peso molecular de um polímero.

Observa-se que na Figura (2-2) os dados estão mais próximos de uma reta, pois o coeficiente de autocorrelação é maior. Note-se que as escalas utilizadas nos dois eixos dos gráficos foram as mesmas.

Quando se calculam vários coeficientes de autocorrelação para vários lags k , em uma mesma seqüência de números, pode-se construir um gráfico de barras com esses diversos coeficientes de

autocorrelação, para se obter importantes informações sobre a estrutura do conjunto de dados e o seu padrão. A este tipo de gráfico dá-se o nome de **autocorrelograma**. Na figura (2-4) damos o autocorrelograma para a seqüência de dados de nosso exemplo do peso molecular.

No gráfico podemos observar que há duas curvas tracejadas acima e abaixo do eixo das abscissas. Estas curvas nos permitem inferir se o coeficiente de correlação para um determinado lag k é significativo ou não. Também estão incluídas neste gráfico, devido ao software utilizado (e em outros autocorrelogramas subseqüentes), valores da estatística t e estatística LBQ (Ljung-Box Q). A estatística t serve para testar se o coeficiente de autocorrelação de um determinado lag é zero ou não, e a estatística LBQ serve para testar se todos os coeficientes de autocorrelação até um determinado lag são iguais a zero ou não.

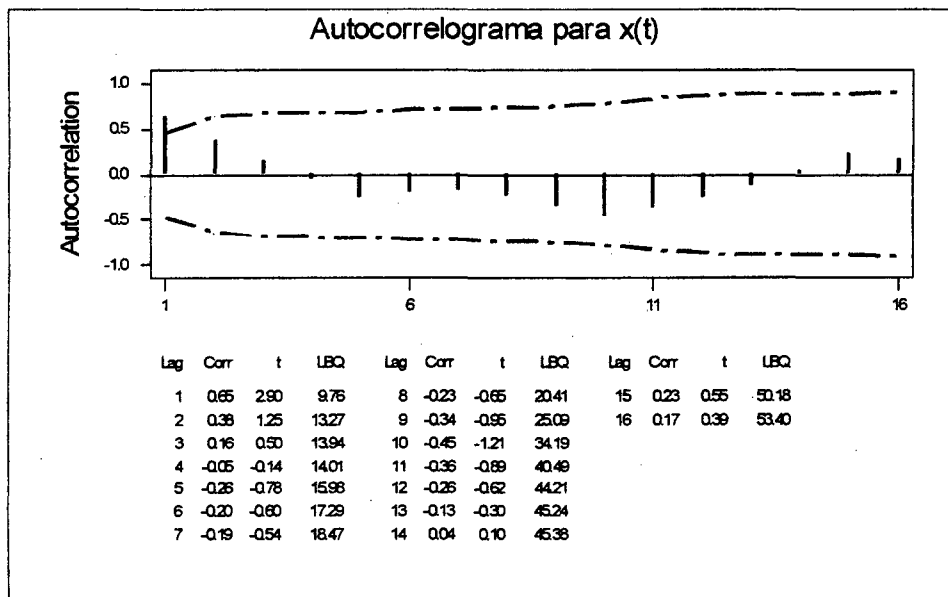


Figura 2-4 – Autocorrelograma para uma variável aleatória (índices de autocorrelação para lags $k=1$ até $k=16$).

2.3-) Influência da Autocorrelação nos Gráficos de Controle Tradicionais

Os gráficos de controle tradicionais (ou de Shewhart) são normalmente robustos em relação a desvios de normalidade nos dados, como vimos no capítulo 1. No entanto, não são robustos se não for válida a suposição de independência estatística das observações sucessivas. Nas indústrias de processos contínuos esta suposição é violada frequentemente, devido à presença de autocorrelação nos dados. Assim sendo, um primeiro efeito da autocorrelação é que as regras de sensibilidade não se aplicam, pois pressupõem a independências das observações.

Vamos mostrar uma outra influência da autocorrelação sobre os gráficos de controle através de um exemplo. A tabela (2-3) traz uma série de 100 valores $y(t)$, gerados por computador (na verdade trata-se de um processo AR(1) com média 0 e $\phi_1 = 0,8$, o qual estudaremos no próximo capítulo).

A figura (2-5) mostra o autocorrelograma para a variável $y(t)$. Observamos que a variável $y(t)$ apresenta níveis significativos de autocorrelação para os lags de 1 a 4. Já na figura (2-6), temos os gráficos de valores individuais e amplitude móvel com limites calculados segundo o método de Shewhart tradicional, baseados na amplitude móvel. Verificamos que há muitos pontos fora dos limites de controle no gráfico de valores individuais. No entanto, eles não eram de se esperar devido ao processo de geração dos valores de $y(t)$.

Esta influência da autocorrelação sobre os gráficos de controle pode ser explicada da seguinte forma: quando o coeficiente de autocorrelação para lag 1 é significativo, as observações sucessivas tendem a ter valores semelhantes, de forma que a amplitude móvel de ordem 2 tenderá a assumir valores pequenos. Deste modo, subestima-se a variabilidade do processo e limites mais estreitos fatalmente irão sinalizar um maior número de alarmes falsos.

t	y (t)	t	y (t)	t	y (t)
1	-0.150	35	0.959	69	0.942
2	-0.759	36	0.725	70	0.998
3	-0.485	37	0.318	71	0.834
4	0.250	38	0.592	72	1.082
5	0.799	39	0.283	73	1.297
6	1.506	40	0.605	74	0.719
7	0.113	41	-0.238	75	0.114
8	-0.027	42	-0.614	76	0.647
9	0.526	43	-1.252	77	-0.083
10	-0.122	44	-1.183	78	-0.846
11	-0.443	45	-0.963	79	-0.321
12	-1.200	46	-0.756	80	0.062
13	-1.883	47	-0.766	81	1.153
14	-1.995	48	0.484	82	1.644
15	-1.983	49	-0.484	83	1.967
16	-2.645	50	-0.755	84	1.630
17	-2.400	51	-1.893	85	1.305
18	-2.122	52	-0.791	86	1.271
19	-1.630	53	-1.272	87	1.004
20	-1.487	54	-1.345	88	0.276
21	-1.353	55	-0.697	89	-0.667
22	-1.268	56	-0.324	90	-0.119
23	-0.343	57	0.178	91	0.127
24	-0.317	58	0.440	92	0.410
25	-0.347	59	-0.334	93	0.435
26	-0.534	60	-0.825	94	-0.165
27	0.559	61	-0.313	95	0.487
28	0.880	62	-0.089	96	0.234
29	1.892	63	-0.541	97	-0.233
30	1.186	64	-0.553	98	-0.597
31	1.780	65	-0.377	99	-0.692
32	0.617	66	-0.023	100	-0.780
33	0.763	67	0.051		
34	1.062	68	-0.414		

Tabela 2-3 –100 valores gerados para a variável $y(t)$ de controle de um processo, modelada por uma série AR(1).

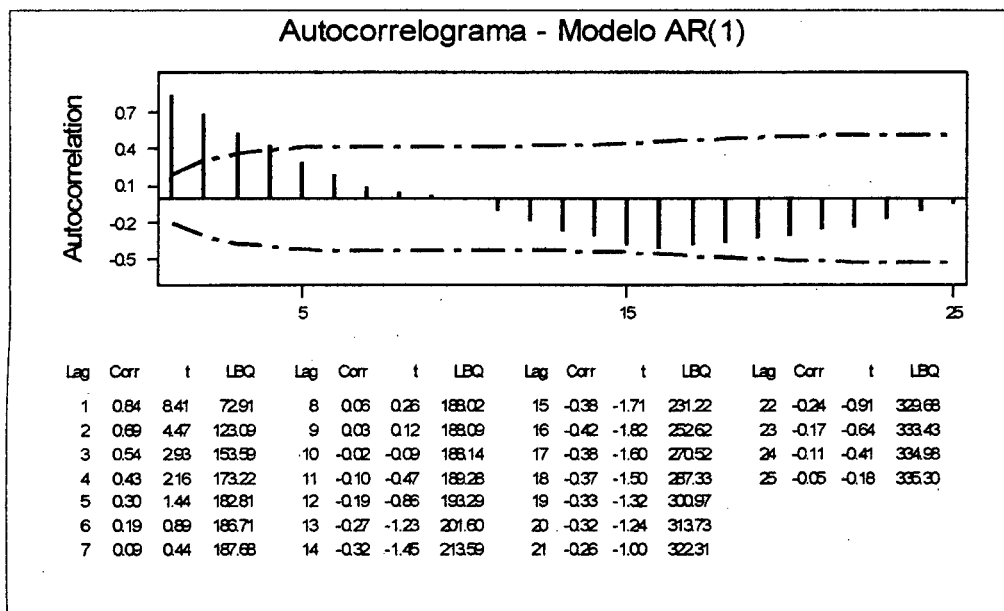


Figura 2-5 – Autocorrelograma para a variável $y(t)$, incluindo os valores discretos para diversos lags.

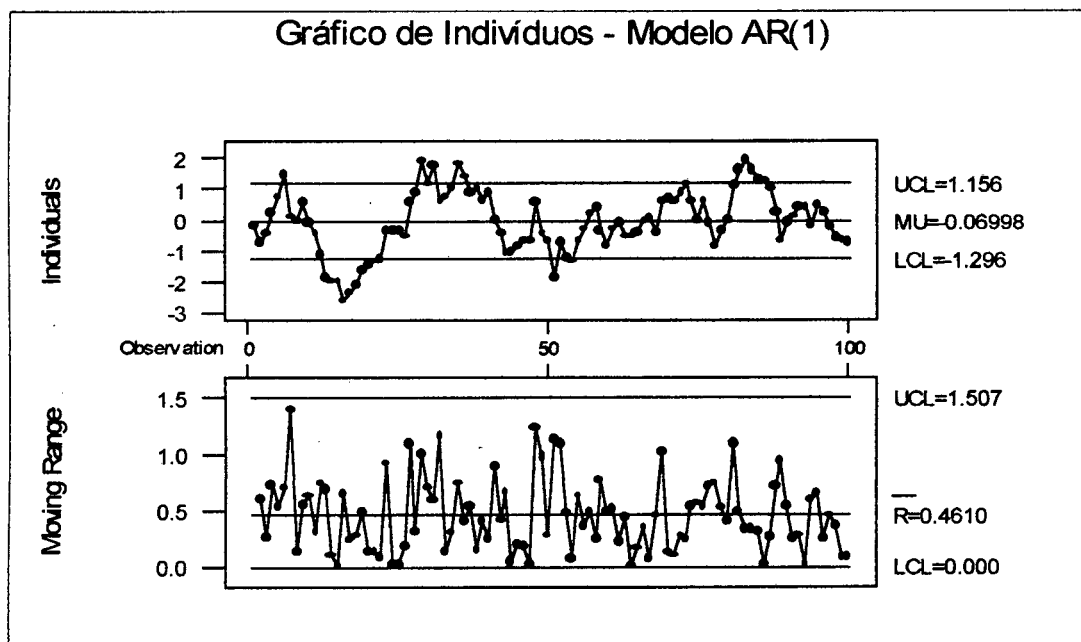


Figura 2-6 – Gráficos de valores individuais e amplitude móvel para $y(t)$, variável gerada segundo o modelo AR(1).

Diversos autores já enfatizaram a influência que a autocorrelação exerce nos gráficos de controle tradicionais, entre eles VASILOPOULOS & STAMBOULIS (1978), ALWAN & ROBERTS (1988), MONTGOMERY & MASTRANGELO (1991) e MARAGAH & WOODALL (1992). De uma forma geral, estes autores mostram que níveis baixos de autocorrelação já causam perturbações nas propriedades desses gráficos, provocando inclusive uma grande quantidade de alarmes falsos.

MARAGAH & WOODALL (1992) fizeram um estudo neste sentido, utilizando séries geradas com base no modelo AR(1), cuja fórmula é dada abaixo, para diversos valores de ϕ_1 e média $\mu = 0$ (sendo os a_t normais e independentes para $t = 1, 2, 3, \dots$):

$$y_t = \phi_1 y_{t-1} + a_t \quad (2-5)$$

Como veremos no Capítulo 3, a correlação entre y_t e y_{t-1} é igual a ϕ_1 . Foram gerados 5000 experimentos de 25 observações cada, para cada valor de ϕ_1 entre $-0,9$ e $0,9$ com passo $0,1$. Para os diferentes valores de ϕ_1 obteve-se o número médio de pontos fora dos limites em cada conjunto de 25 pontos, o que é mostrado na tabela (2-4).

Como se verifica na tabela (2-4), o número médio de pontos fora dos limites de controle é mínimo para $\phi_1 = -0,9$, crescendo e chegando a um máximo quando $\phi_1 = 0,9$. Tomando a probabilidade α do erro tipo I para o gráfico de Shewhart tradicional com distribuição normal temos $\alpha = 0,27\%$. Desta forma o número esperado de pontos fora dos limites para o gráfico tradicional sob controle com $n = 25$ é $25 \cdot (0,0027) = 0,067$. Comparando com os valores da tabela (2-4), verificamos que este valor é próximo daquele obtido para $\phi_1 = 0$ ($0,094$). Para $\phi_1 > 0$ o número de alarmes falsos é maior que o dos gráficos tradicionais, enquanto que para $\phi_1 < 0$ o número de alarmes falsos é menor que o dos gráficos tradicionais. Notamos que a situação $\phi_1 < 0$ não é usual na indústria.

O número pequeno de alarmes falsos para valores negativos de ϕ_1 é na verdade uma consequência do aumento do intervalo entre os limites de controle, devido à superestimação da variabilidade do processo (com a autocorrelação negativa a distância entre valores sucessivos tende a ser maior, levando a um valores maiores para a amplitude móvel e para a estimativa da

variabilidade). Esta é também uma condição perigosa pois o gráfico de CEP se torna insensível a causas especiais de variação.

$\phi_1 = \rho$	Número médio de pontos fora dos limites
-0,9	0,001
-0,8	0,002
-0,7	0,004
-0,6	0,007
-0,5	0,011
-0,4	0,017
-0,3	0,027
-0,2	0,036
-0,1	0,058
0	0,094
0,1	0,135
0,2	0,202
0,3	0,312
0,4	0,498
0,5	0,752
0,6	1,138
0,7	1,807
0,8	2,738
0,9	4,081

Tabela 2-4 – Número médio de pontos fora dos limites de controle para um modelo AR(1) com $n = 25$.

No próximo capítulo abordamos algumas soluções para tratar o problema da autocorrelação em gráficos de controle.

Capítulo 3 – Soluções de CEP em Meios Contínuos com Dados Autocorrelacionados

Conforme já vimos anteriormente, em processos contínuos há geralmente a presença de autocorrelação nos dados a serem analisados por ferramentas de CEP. A autocorrelação nos dados prejudica a análise e pode induzir a conclusões errôneas sobre o processo. Entre as técnicas utilizadas para contornar o problema da autocorrelação estão o uso de **Modelos Autorregressivos**, (chamados também de $AR(k)$), uma solução baseada no gráfico **EWMA** e a **solução de Wheeler**.

Neste capítulo expomos os detalhes das soluções acima, que são utilizadas de forma alternativa ao gráfico de Shewhart tradicional para CEP com dados autocorrelacionados.

3.1-) Modelos Autorregressivos ($AR(k)$) e sua Aplicação ao Controle de Processos Contínuos.

Nesta seção conceituamos os modelos $AR(k)$, dando as suas principais características e condições para uso. Em seguida descrevemos suas aplicações no caso do CEP com dados autocorrelacionados. As referências para esta seção são VANDAELE (1983) e MAKRIDAKIS (1983).

3.1.1-) Conceito de Modelos Autorregressivos ($AR(k)$)

Os modelos autorregressivos ($AR(k)$) são utilizados para descrever **séries temporais** (seqüências de observações coletadas ao longo do tempo). Sendo **modelos**, procuram explicar o comportamento de uma série de valores observados experimentalmente.

É possível calcular os desvios em relação a esse modelo, os quais possibilitam avaliar se o modelo descreve adequadamente a situação real.

O modelo $AR(k)$ pressupõe que as observações da série sejam variáveis aleatórias, todas com a mesma média e mesma variância. A isto damos o nome de **condição de estacionariedade**. Supõe também que a observação atual depende de observações anteriores, ou seja, que há autocorrelação. Esta autocorrelação é mais forte para observações adjacentes, diminuindo conforme

as observações vão se distanciando. A condição de estacionariedade, como veremos a seguir, ainda nos garante que as autocorrelações já se tornam pequenas para “lags” não muito grandes, da ordem de 4 ou 5. Estas são as principais características do modelo AR(k).

O caso mais simples do modelo AR(k) se dá para $k = 1$, ou seja, AR(1). Este modelo é chamado de **modelo autorregressivo de 1ª ordem**. Neste modelo, o valor da observação atual pode ser expresso como uma função linear do valor prévio da série e uma variável aleatória a_t .

$$x_t = (1 - \phi_1) \mu + \phi_1 x_{t-1} + a_t \quad (3-1)$$

Onde:

- x_t = Valor da observação atual
- x_{t-1} = Valor da observação anterior
- ϕ_1 = Parâmetro (a ser estimado)
- μ = Média do processo
- a_t = Variável aleatória

As variáveis aleatórias a_t ($t = 1, 2, 3, \dots$) são normais e independentemente distribuídas com média 0, variância fixa para todo t e não correlacionadas com x_{t-1} .

A expressão (3-1) acima nos indica, portanto, uma dependência direta de x_t com x_{t-1} , ou seja, os valores atuais do processo dependem dos valores imediatamente anteriores.

A condição de estacionariedade, que já mencionamos acima, impõe a seguinte restrição para que o processo tenha uma variância constante e finita ao longo do tempo:

$$|\phi_1| < 1 \quad (3-2)$$

O coeficiente de autocorrelação para a variável Y em função do lag k será dada por:

$$\rho_k = \phi_1^k \quad (3-3)$$

Sendo $|\phi_1| < 1$, concluímos que os coeficientes de autocorrelação decrescem conforme aumenta o lag k.

Finalmente, a variância de x_t para o modelo AR(1) é dada por:

$$\sigma^2(x) = \frac{\sigma_a^2}{1 - \phi^2} \quad (3-4)$$

Há também modelos autorregressivos de ordens maiores que 1. Em diversas aplicações, pode ocorrer a influência de observações mais defasadas em relação a Y_t . Por exemplo, se os eventos que ocorrerem há dois períodos irão influenciar o comportamento atual, então podemos incluir Y_{t-2} no modelo. Dessa forma temos:

$$x_t = \delta_2 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t \quad (3-5)$$

onde ϕ_1 e ϕ_2 são parâmetros a serem estimados. A equação (3-5) acima representa o modelo AR(2). Sendo μ a média do processo, a constante δ_2 acima tem o valor,:

$$\delta_2 = (1 - \phi_1 - \phi_2) \mu \quad (3-6)$$

As condições de estacionariedade, os coeficientes de autocorrelação e a variância para o AR(2) são dadas abaixo:

$$\text{Condições de estacionariedade: } \phi_1 + \phi_2 < 1; \phi_1 - \phi_2 < 1; |\phi_2| < 1 \quad (3-7)$$

$$\text{Coeficientes de autocorrelação: } \rho_0 = 1; \rho_1 = \phi_1 / (1 - \phi_2); \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, k > 1 \quad (3-8)$$

$$\text{Variância: } \sigma^2(x) = \frac{\sigma_a^2(1 - \phi_2)}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \quad (3-9)$$

Analogamente, para incluirmos a influência de k observações anteriores a x_t , teríamos o modelo AR(k):

$$x_t = \delta_k + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_k x_{t-k} + a_t \quad (3-10)$$

As condições de estacionariedade, média, variância e coeficientes de autocorrelação com lag k podem também ser obtidos para os modelos $AR(k)$, como no caso do $AR(1)$ e $AR(2)$, porém as fórmulas são bem mais complexas.

3.1.2-) Aplicação de Modelos Autorregressivos ($AR(k)$) ao Controle de Processos

Uma das possíveis soluções possíveis para tratar o problema da autocorrelação em gráficos de controle é o uso de modelos de séries temporais (ALWAN & ROBERTS (1988)) para eliminar ou minimizar os efeitos da autocorrelação. Estes modelos incluem como um caso especial o modelo autorregressivo $AR(k)$. Damos a seguir um exemplo da aplicação dos modelos que estamos estudando para contornar este problema.

Exemplo de Aplicação 3.1

Consideremos os dados, apresentados na tabela (3-1), do peso molecular de um polímero, tomados a cada 2 horas. Da função autocorrelação (autocorrelograma) da figura (3-1) obtemos um valor significativo para o r_1 (0,66) e concluímos que o processo apresenta dados autocorrelacionados. Analisando o gráfico de valores individuais de Shewhart na figura (3-2), com o desvio padrão estimado utilizando a amplitude móvel, observamos que há vários pontos fora dos limites de controle que indicariam causas especiais de variação no processo. Porém, não nos sentimos em condições de fazer tal afirmação, devido à presença de autocorrelação nos dados.

t	x(t)	t	x(t)	t	x(t)
1	2048	26	2002	51	2016
2	2025	27	2010	52	2019
3	2017	28	1975	53	2036
4	1995	29	1983	54	2015
5	1983	30	2021	55	2032
6	1943	31	2051	56	2016
7	1940	32	2056	57	2000
8	1947	33	2018	58	1988
9	1972	34	2030	59	2010
10	1983	35	2023	60	2015
11	1935	36	2036	61	2029
12	1948	37	2019	62	2019
13	1966	38	2000	63	2016
14	1954	39	1986	64	2010
15	1970	40	1952	65	2006
16	2039	41	1988	66	2009
17	2015	42	2016	67	1990
18	2021	43	2002	68	1986
19	2010	44	2004	69	1947
20	2012	45	2018	70	1958
21	2003	46	2002	71	1983
22	1979	47	1967	72	2010
23	2006	48	1994	73	2000
24	2042	49	2001	74	2015
25	2000	50	2013	75	2032

Tabela 3-1 – Medidas do peso molecular de um polímero, feitas a cada 2 horas, em função do número da observação t (fonte: MONTGOMERY (1996, pág. 426))

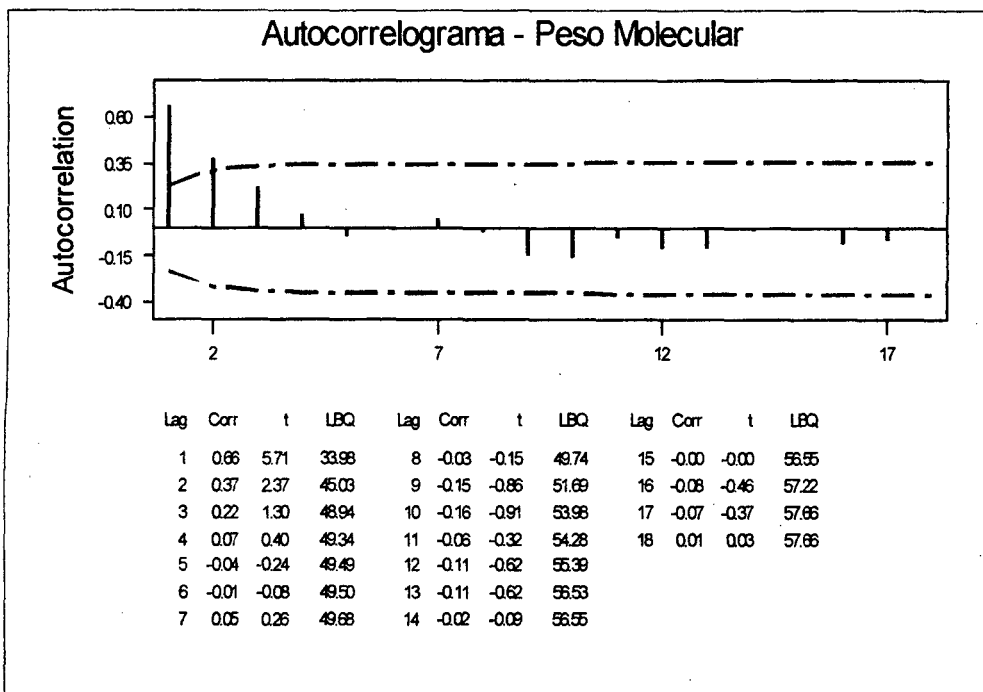


Figura 3-1 – Autocorrelograma para a seqüência de dados do peso molecular de um polímero.

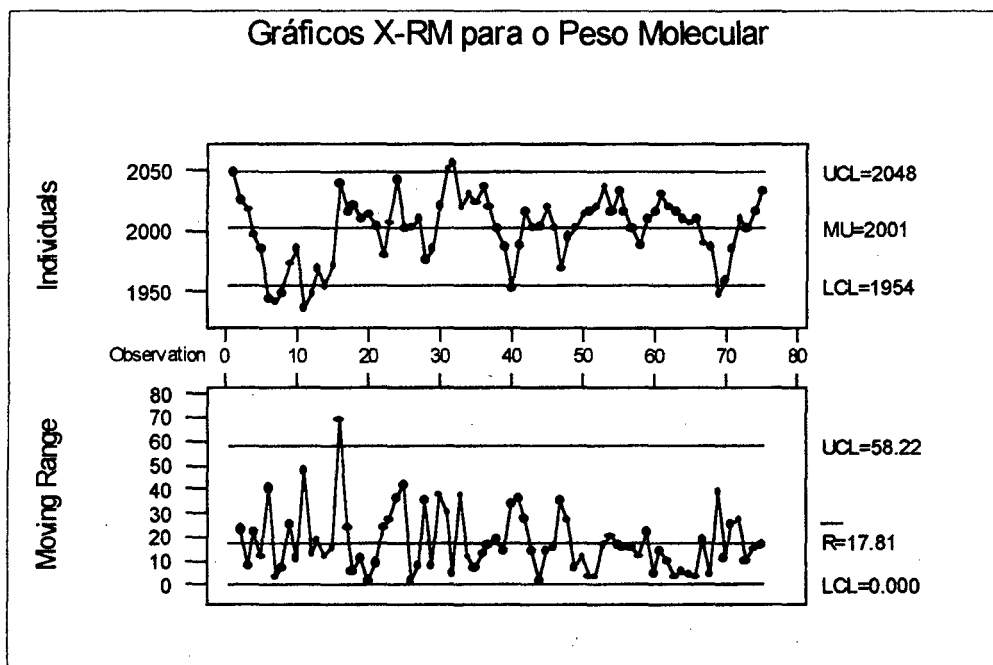


Figura 3-2 – Gráfico de valores individuais e amplitude móvel (X-RM) para o peso molecular de um polímero (valores observados a cada 2 horas).

Passamos então a modelar a autocorrelação através de um modelo autorregressivo AR(k) que descreva o mais adequadamente possível a variação dos dados do processo. A aplicação desta técnica requer, segundo NEWBOLD (1995):

- A estimação de parâmetros do modelo: pode ser feita utilizando-se o método dos mínimos quadrados.
- A seleção da ordem k: utiliza-se o método chamado "forward selection". Experimentalmente, podemos calcular modelos de diversas ordens (1, 2, 3, etc.) gradativamente e utilizar o seguinte testes de hipóteses para verificar se o coeficiente da ordem mais alta do modelo (ϕ_k) é igual a zero ($H_0 : \phi_k = 0$ contra $H_1 : \phi_k \neq 0$), sendo Z uma variável normal padrão e $P(Z > Z_{\alpha/2}) = \alpha/2$:

$$\text{Rejeitar } H_0 \text{ se } \frac{\hat{\phi}_k}{s_k} < -Z_{\alpha/2} \quad \text{ou} \quad \frac{\hat{\phi}_k}{s_k} > Z_{\alpha/2} \quad (3-11)$$

- Uma "regra prática" é rejeitar H_0 se a estimativa $\hat{\phi}_k$ tiver valor em módulo maior que duas vezes seu erro padrão s_k (o que equivale a adotar $\alpha \sim 5\%$). Este processo é repetido até que se encontre um modelo para o qual a hipótese $\phi_k = 0$ não seja rejeitada e então adota-se o modelo de ordem imediatamente anterior. Segundo MONTGOMERY & RUNGER (1994), é bastante incomum na prática encontrar modelos autorregressivos de ordem maior do que dois.

A validação do modelo obtido se faz pela análise dos resíduos. Estes são dados pela diferença entre o valor real e o previsto no modelo, conforme (3-12). Se o modelo for adequado para representar a variação dos dados do processo, os resíduos \hat{a}_t serão IID (independentes e identicamente distribuídos), o que podemos testar usando o autocorrelograma dos mesmos.

$$\hat{a}_t = x_t - \hat{x}_t \quad (3-12)$$

Sendo os resíduos não autocorrelacionados, podem ser analisados com gráficos de Shewhart tradicionais e uma evidência de falta de controle apontada por esses gráficos será tomada como sugestão de provável causa especial.

No nosso caso o modelo AR(1) obtido com o software Minitab foi o seguinte:

$$\hat{x}_t = 602,159 + 0,699 x_{t-1} \quad (s_1 = 0,0851)$$

O erro padrão para $\hat{\phi}_1$, s_1 , obtido foi 0,0851. Pela “regra prática” acima, rejeitamos a hipótese deste coeficiente ser igual a zero.

Para o modelo AR(2), obtivemos o seguinte:

$$\hat{x}_t = 690,273 + 0,789 x_{t-1} - 0,133 x_{t-2} \quad (s_2 = 0,117)$$

Neste caso, o coeficiente $\hat{\phi}_2$ não tem um valor em módulo maior que duas vezes seu erro padrão, então aceitamos a hipótese de que este coeficiente é igual a zero e adotamos o modelo anterior AR(1).

Através do autocorrelograma dos resíduos do AR(1), dado na figura (3-3), verificamos que os mesmos não são correlacionados pois estão dentro dos limites de aceitação. Na figura (3-4) temos o gráfico X-RM para estes resíduos. Observamos que agora aparece um ponto fora dos limites no gráfico de RM correspondente à observação 17, indicando possível causa especial de variação no processo.

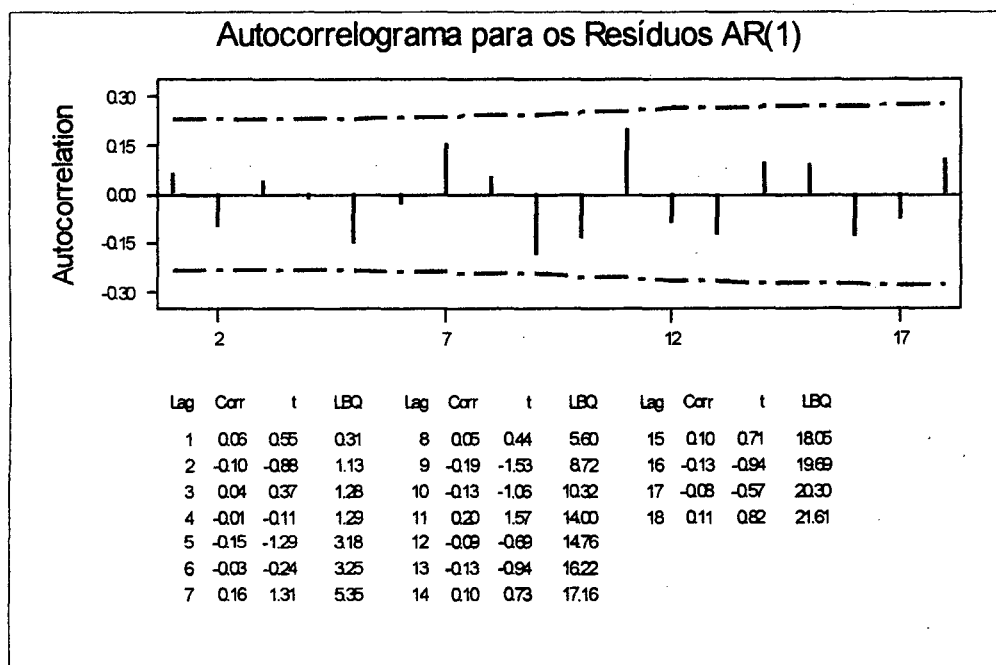


Figura 3-3 – Autocorrelograma dos resíduos do modelo AR(1).

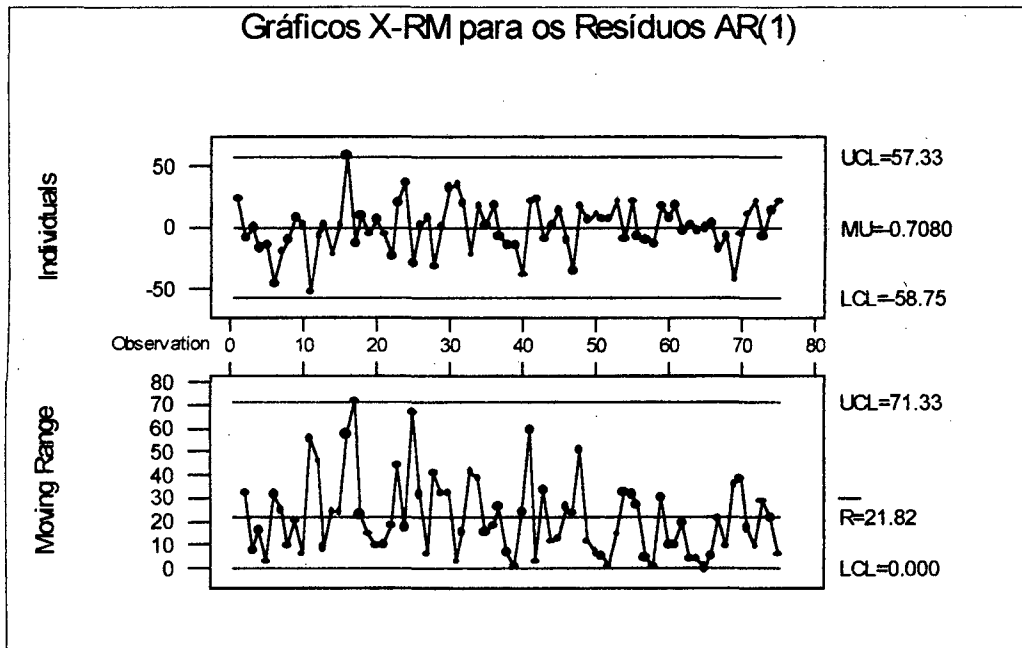


Figura 3-4 – Gráficos X-RM para os resíduos após aplicação do modelo AR(1).

Exemplo de Aplicação 3.2

A viscosidade de um produto químico, em centistokes, foi tabulada a cada 2 minutos, conforme a tabela (3-2) abaixo.

t	y(t)	t	y(t)	t	y(t)	t	y(t)
1	29.330	26	33.220	51	27.990	76	24.280
2	19.980	27	30.150	52	24.130	77	22.690
3	25.760	28	27.080	53	29.200	78	26.600
4	29.000	29	33.660	54	34.300	79	28.860
5	31.030	30	36.580	55	26.410	80	28.270
6	32.680	31	29.040	56	28.780	81	28.170
7	33.560	32	28.080	57	21.280	82	28.580
8	27.500	33	30.280	58	21.710	83	30.760
9	26.750	34	29.350	59	21.470	84	30.620
10	30.550	35	33.600	60	24.710	85	20.840
11	28.940	36	30.290	61	33.610	86	16.560
12	28.500	37	20.110	62	36.540	87	25.230
13	28.190	38	17.510	63	35.700	88	31.790
14	26.130	39	23.710	64	33.680	89	32.520
15	27.790	40	24.220	65	29.290	90	30.280
16	27.630	41	32.430	66	25.120	91	26.140
17	29.890	42	32.440	67	27.230	92	19.030
18	28.180	43	29.390	68	30.610	93	24.340
19	26.650	44	23.450	69	29.060	94	31.530
20	30.010	45	23.620	70	28.480	95	31.950
21	30.800	46	28.120	71	32.010	96	31.680
22	30.450	47	29.940	72	31.890	97	29.100
23	36.610	48	30.560	73	31.720	98	23.150
24	31.400	49	32.300	74	29.090	99	26.740
25	30.830	50	31.580	75	31.920	100	32.440

Tabela 3-2 – Viscosidade de um produto químico. (Fonte: MONTGOMERY, 1996, pág. 428).

O autocorrelograma e os gráficos de valores individuais e amplitude móvel para estes dados podem ser visto nas figuras (3-5) e (3-6), nas quais observamos uma clara presença de autocorrelação no processo, ocasionando vários pontos fora dos limites de controle no gráfico de valores individuais. Seguindo o mesmo procedimento descrito no exemplo anterior, foi obtido com o auxílio do software Minitab um modelo AR(2) para este processo, cuja fórmula é a seguinte:

$$\hat{x}_t = 20,502 + 0,7193 x_{t-1} - 0,4349 x_{t-2}$$

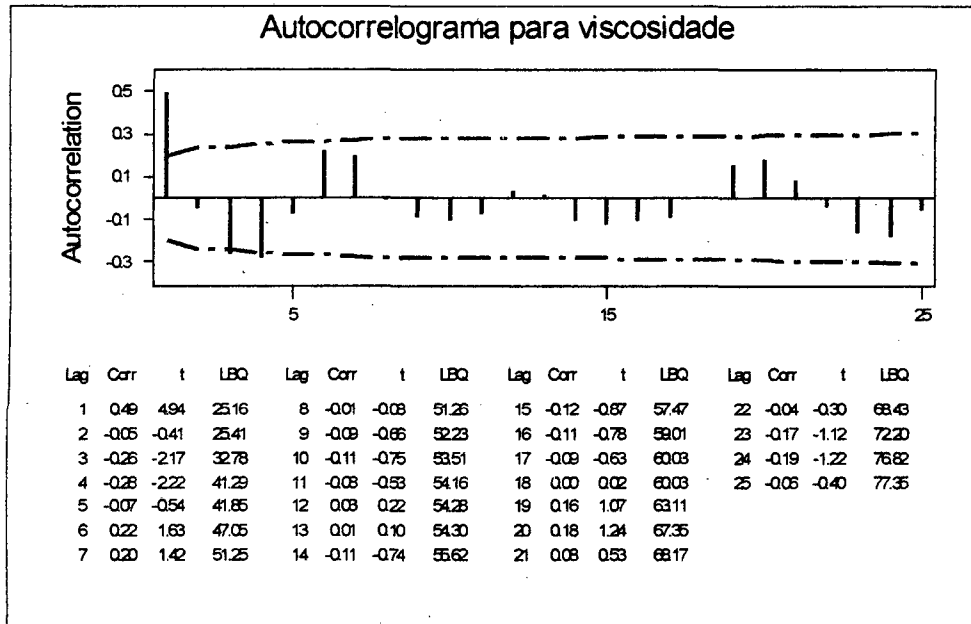


Figura 3-5 – Autocorrelograma para a viscosidade de um produto químico.

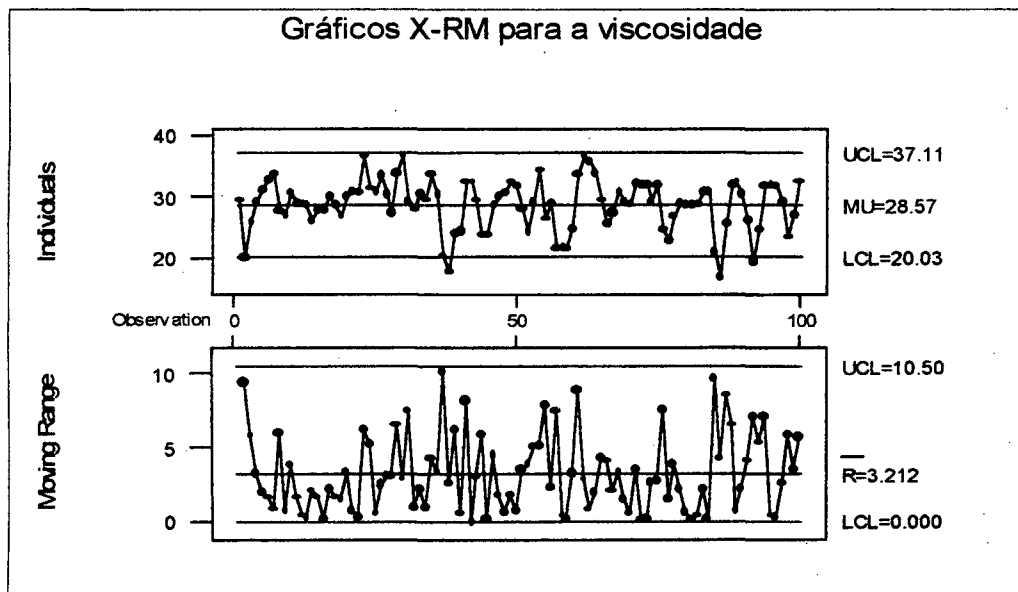


Figura 3-6 – Gráficos X-RM para a viscosidade de um produto químico (com a presença de autocorrelação).

Nas figuras (3-7) e (3-8) damos, finalmente, o autocorrelograma e gráficos X-RM para os resíduos do modelo AR(2) aplicado. Os resíduos não são correlacionados. O gráfico RM apresenta um ponto fora dos limites de controle, o que sinaliza uma provável causa especial.

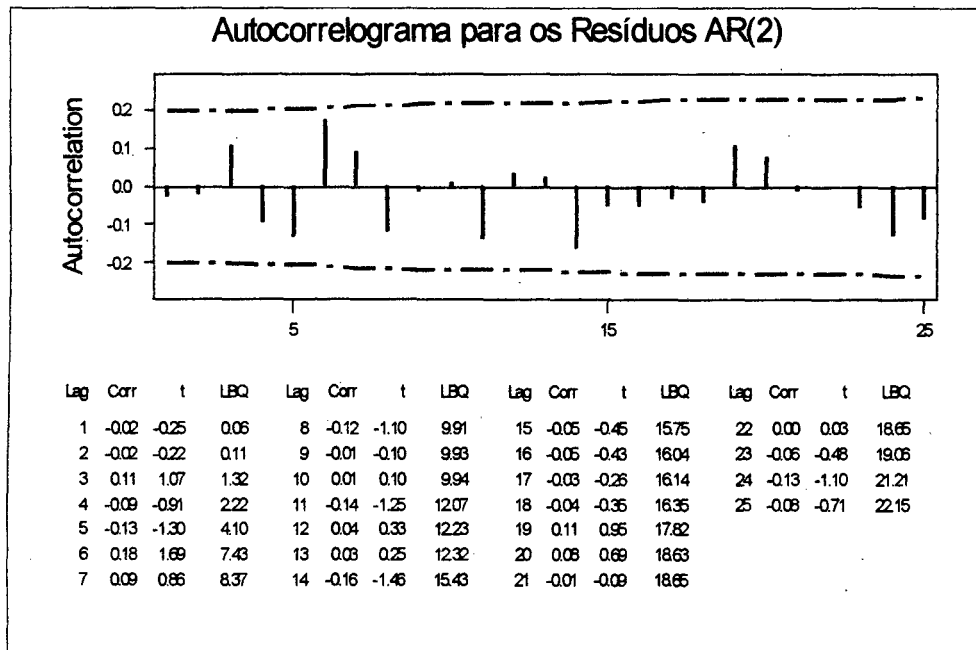


Figura 3-7 – Autocorrelograma para os resíduos do modelo AR(2).

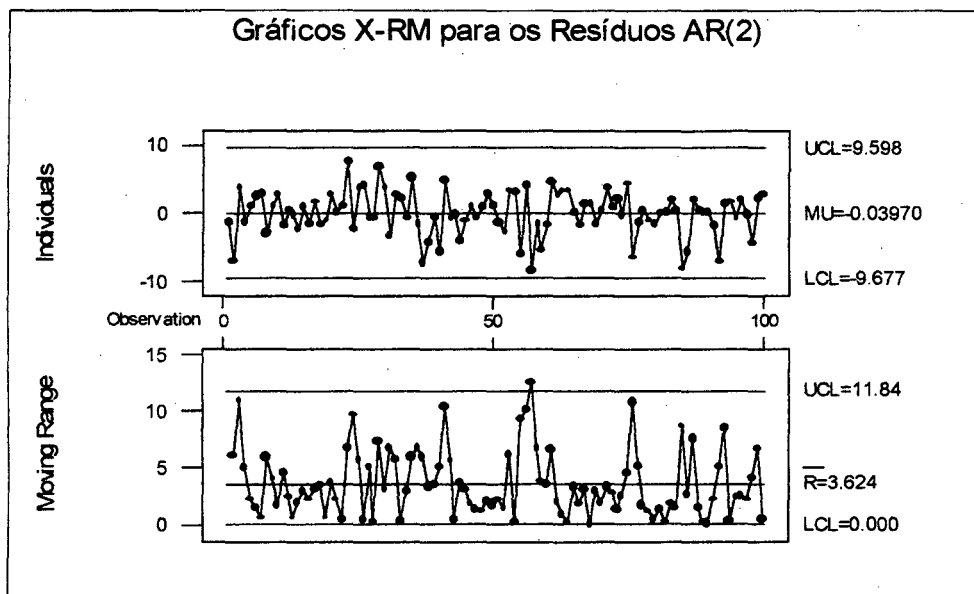


Figura 3-8 – Gráficos X-RM para os resíduos do modelo AR(2).

3.2-)EWMA (“Exponentially Weighted Moving Average”)

3.2.1-) O Gráfico de Controle EWMA

Entre as soluções de CEP para processos contínuos, existe uma que envolve o **gráfico EWMA** (“Exponentially Weighted Moving Average” – “Média Móvel Ponderada Exponencialmente”). Nesta seção damos uma idéia da base teórica do EWMA, que nos será útil ao abordarmos a solução do EWMA para CEP aplicado a dados autocorrelacionados.

O gráfico de controle EWMA foi introduzido pela primeira vez por ROBERTS (1959). Também foi abordado teoricamente por HUNTER (1986), CROWDER (1989) e LUCAS & SACCUCCI (1990), entre outros.

O gráfico de controle EWMA baseia-se na **estatística EWMA**, que é definida como segue:

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}, t = 1, 2, 3, \dots \quad (3-13)$$

onde z_t é o valor da estatística EWMA, x_t é a medida no instante t (médias amostrais ou valores individuais) de um processo e λ ($0 < \lambda < 1$) é uma constante (“peso”). O valor inicial z_0 pode ser o “alvo” (valor desejado) para o processo ou uma média de dados preliminares.

Substituindo-se as expressões de z_{t-1} , z_{t-2} , ..., z_1 recursivamente em (3-13) concluímos que a estatística EWMA é uma média ponderada de todas as observações anteriores, com pesos que decrescem exponencialmente conforme é mais antiga a observação, como vemos em (3-14).

$$z_t = \lambda \sum_{k=0}^{t-1} (1 - \lambda)^k x_{t-k} + (1 - \lambda)^t z_0 \quad (3-14)$$

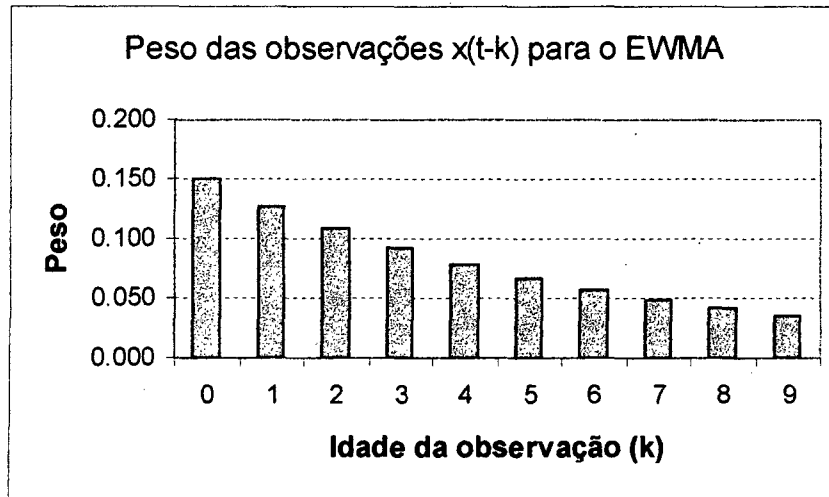


Figura 3-9 - Exemplo de pesos das observações anteriores ao instante t utilizadas no cálculo da estatística EWMA (para $\lambda = 0,15$), em função da idade k da observação.

Mostra-se a partir de (3-14) que a soma dos pesos das observações do processo para formar a EWMA é 1 (ver nota¹). O cálculo da EWMA sempre dá um peso maior às observações mais recentes, conforme observamos na figura (3-9).

O EWMA utiliza os dados históricos de uma maneira bastante adequada. Por exemplo, no gráfico de Shewhart só têm peso as observações (X ou \bar{X}) mais recentes. Já no **gráfico de controle CUSUM** (“Cumulative Sum” – “Soma Cumulativa”), descrito por DUNCAN(1985), são dados pesos iguais a todas as observações anteriores. Dessa forma, segundo HUNTER (1986), o EWMA é uma solução de compromisso entre os gráficos de \bar{X} e CUSUM.

3.2.2-) Aplicação do Gráfico EWMA com Dados Autocorrelacionados

Esta seção está baseada no artigo de MONTGOMERY & MASTRANGELO (1991), que sugere uma solução para o CEP com autocorrelação. Os autores utilizam o fato de que o EWMA pode ser aplicado em certas situações nas quais os dados são autocorrelacionados. Consideremos o caso em que as observações do processo podem ser modeladas pela seguinte equação:

¹ Calculando-se a soma dos termos de uma progressão geométrica.

$$x_t = x_{t-1} + a_t - \theta a_{t-1} \quad (3-15)$$

sendo x_t a medida no instante t , θ um parâmetro do processo e os a_t IID (independentes e identicamente distribuídos), com média 0 e desvio padrão σ_a .

O modelo da equação (3-15) é conhecido como **modelo de média móvel integrada de 1ª ordem**, descrito em VANDAELE (1983). Apresenta um comportamento não estacionário (ou seja, a variável x_t “flutua” como se não houvesse um valor fixo para a média do processo). Este modelo pode ocorrer em fábricas da área química e de processos contínuos, quando x_t é uma variável de saída do processo “não controlado”, ou seja, nenhuma ação de controle é tomada para manter a variável próxima a um valor alvo.

Consideremos agora a equação da estatística EWMA já dada em (3-13):

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1} \quad (3-16)$$

Segundo os autores, a estatística EWMA acima, com $\lambda = 1 - \theta$, é um previsor ótimo para a observação imediatamente posterior de x_t . Pode-se ver uma demonstração em VANDAELE (1983, pág. 147-8). Isto significa que a previsão no instante t , para a observação x_{t+1} , será dada por z_t , ou seja:

$$\hat{x}_{t+1}(t) = z_t \quad (3-17)$$

Se o processo seguir o modelo da equação (3-15), então os autores afirmam que os erros de previsão e_t para a observação imediatamente posterior (dados pela equação (3-18)) serão independentemente e identicamente distribuídos. O parâmetro λ é escolhido de forma a minimizar a soma dos quadrados desses erros de previsão.

$$e_t = x_t - \hat{x}_t(t-1) \quad (3-18)$$

Portanto, se após analisarmos os erros de previsão (também chamados de resíduos do modelo), verificarmos que não apresentam autocorrelação, podemos controlar estes erros utilizando gráficos de controle de Shewhart usuais.

Suponhamos agora que o processo não siga exatamente o modelo de (3-18). De acordo com os autores do artigo, se as observações do processo forem positivamente correlacionadas e a média do processo não flutuar “rapidamente”, o EWMA com um valor apropriado de λ (escolhido segundo o mesmo critério acima) ainda fornecerá um bom previsor para a observação imediatamente posterior. Nestas condições, o gráfico EWMA torna-se uma aproximação do modelo exato de séries temporais para o processo.

O gráfico dos resíduos do EWMA, apesar de fornecer informações sobre o estado de controle do processo, não dá uma visão da dinâmica do processo ao longo do tempo. Para tanto, é necessário construir um outro gráfico com as observações x_{t+1} e os correspondentes valores da estatística z_t superpostos. MONTGOMERY & MASTRANGELO (1991) mostram que é possível combinar informações sobre o estado de controle do processo e dinâmica do processo em um só gráfico. Este gráfico chama-se **EWMA com linha média móvel**. É de construção mais complexa que o gráfico de resíduos e não será abordado neste trabalho.

Exemplo de aplicação 3.3

Vamos apresentar um exemplo de aplicação do gráfico EWMA ao CEP com dados autocorrelacionados. Os dados da tabela (3-3) são leituras de concentração de um processo químico, em g/l, feitas a cada 30 minutos. Observa-se na figura (3-10) que o coeficiente de autocorrelação para o lag 1 é significativo (0,75). Portanto, a autocorrelação nos dados deve ser levada em conta.

O gráfico RM da figura (3-11) mostra um ponto fora dos limites de controle, porém não podemos afirmar que há falta de controle estatístico devido à presença de autocorrelação.

t	x(t)	t	x(t)	t	x(t)	t	x(t)	t	x(t)
1	204	21	190	41	208	61	207	81	200
2	202	22	196	42	209	62	204	82	202
3	201	23	199	43	209	63	201	83	202
4	202	24	203	44	206	64	197	84	207
5	197	25	199	45	200	65	189	85	206
6	201	26	207	46	203	66	189	86	211
7	198	27	204	47	202	67	196	87	205
8	188	28	207	48	195	68	193	88	210
9	195	29	209	49	196	69	193	89	210
10	189	30	205	50	203	70	198	90	198
11	195	31	202	51	196	71	194	91	194
12	192	32	200	52	197	72	198	92	192
13	196	33	208	53	197	73	199	93	189
14	194	34	214	54	203	74	204	94	188
15	196	35	205	55	205	75	200	95	189
16	199	36	211	56	194	76	203	96	194
17	197	37	212	57	199	77	200	97	194
18	197	38	214	58	201	78	197	98	198
19	192	39	210	59	198	79	196	99	196
20	195	40	208	60	202	80	202	100	200

Tabela 3-3 – Valores da concentração de um processo químico. (fonte: MONTGOMERY, 1996, pg. 426)

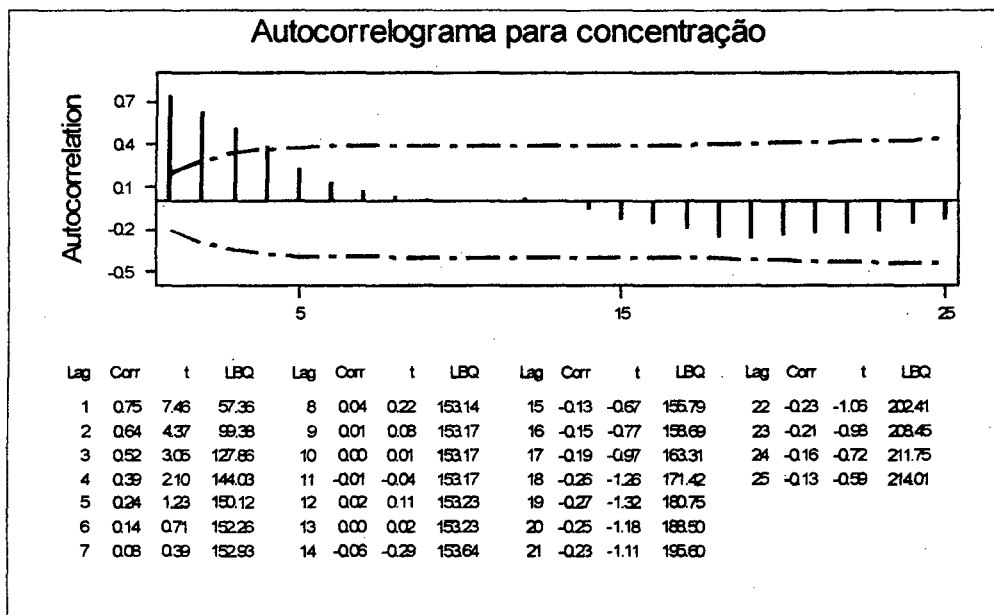


Figura 3-10 – Autocorrelograma para as observações de concentração de um produto químico.

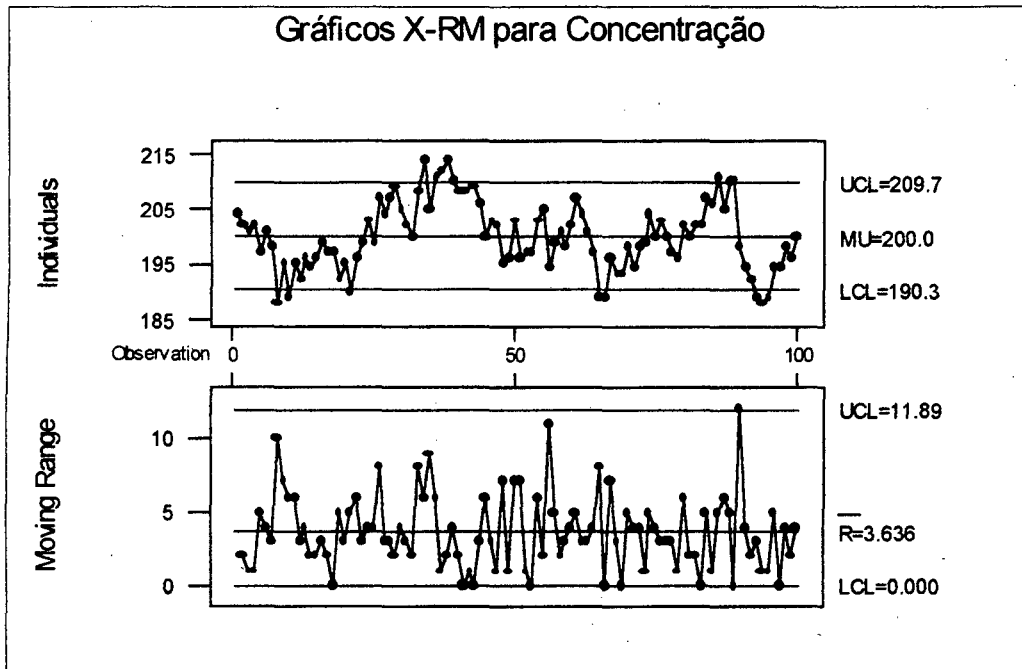


Figura 3-11 – Gráficos X-RM para a concentração de um produto químico.

Vamos aplicar a solução EWMA para estes dados. Calculamos as estatísticas EWMA e respectivos erros quadráticos de previsão, de acordo com a equação:

$$e_t^2 = (x_t - z_{t-1})^2$$

Utilizamos valores de λ de 0,1 a 0,9 com passo de 0,1, obtivemos $\lambda = 0,7$ como o valor que minimizou a soma dos erros quadráticos e_t^2 . A análise dos resíduos obtidos para este valor de λ mostrou que eles não estavam correlacionados, conforme o autocorrelograma da figura (3-12).

Passamos então a analisar os gráficos X-RM para os resíduos, figura (3-13). O gráfico de amplitude móvel indica condição de controle. Já o gráfico de valores individuais mostra uma seqüência de pontos ascendentes após a observação 90, indicando uma possível existência de causa especial.

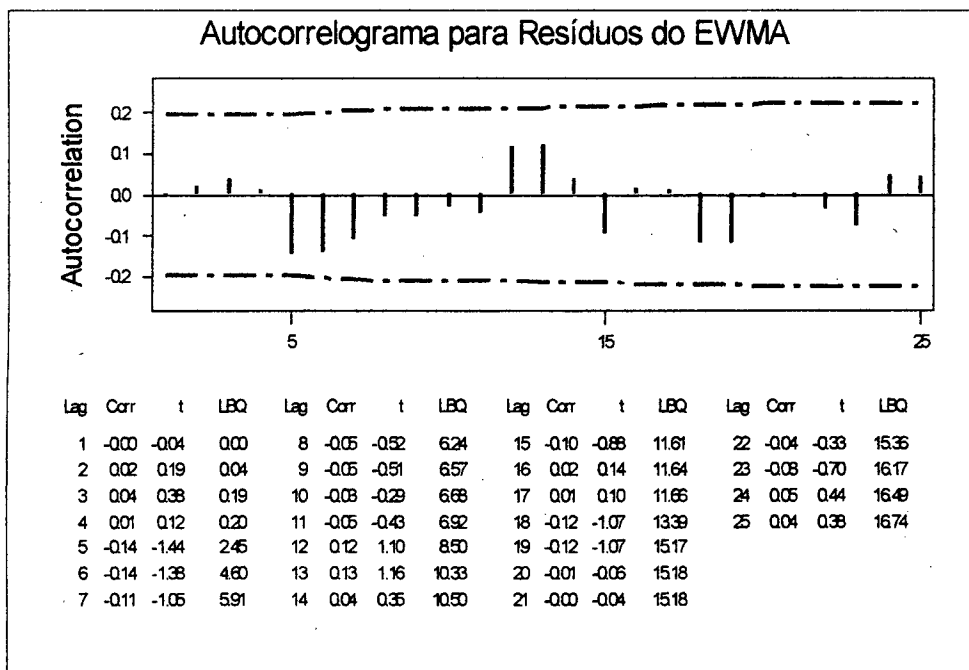


Figura 3-12 – Autocorrelograma para os resíduos (erros de previsão) do EWMA com $\lambda = 0,7$.

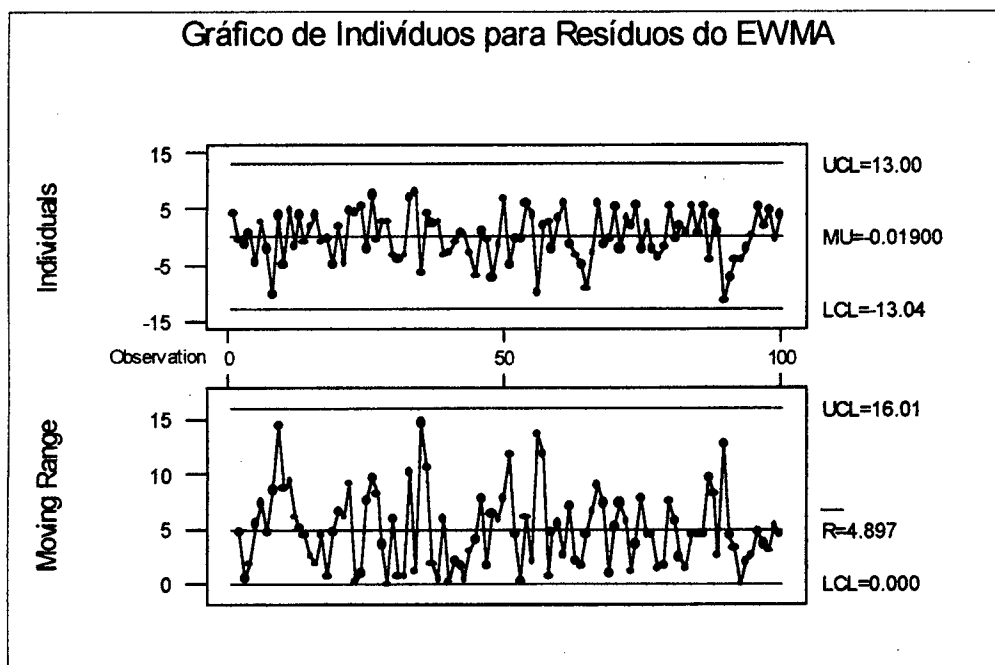


Figura 3-13 – Gráficos X-RM para os resíduos (erros de previsão) do EWMA com $\lambda = 0,7$ e limites de Shewhart.

Exemplo de aplicação 3.4

Consideremos novamente o exemplo de aplicação 3.1 para o peso molecular de um polímero. Através de análise dos gráficos X-RM dos resíduos do modelo AR(1) dados na figura (3-4), página 59, havíamos observado a existência de um ponto fora dos limites de controle do gráfico RM, correspondente à observação 17.

Utilizando a solução do EWMA com autocorrelação neste mesmo exemplo, obtivemos o valor de $\lambda = 0,9$ (minimizando a somatória dos erros de previsão quadráticos). Os resíduos calculados não se apresentaram correlacionados, como se vê no autocorrelograma da figura (3-14).

Os gráficos X-RM para os resíduos do EWMA são dados na figura (3-15). Verificamos que o gráfico da amplitude móvel mostra um ponto fora dos limites correspondente à observação 17, o que é equivalente ao resultado obtido no caso do gráfico para resíduos do AR(1).

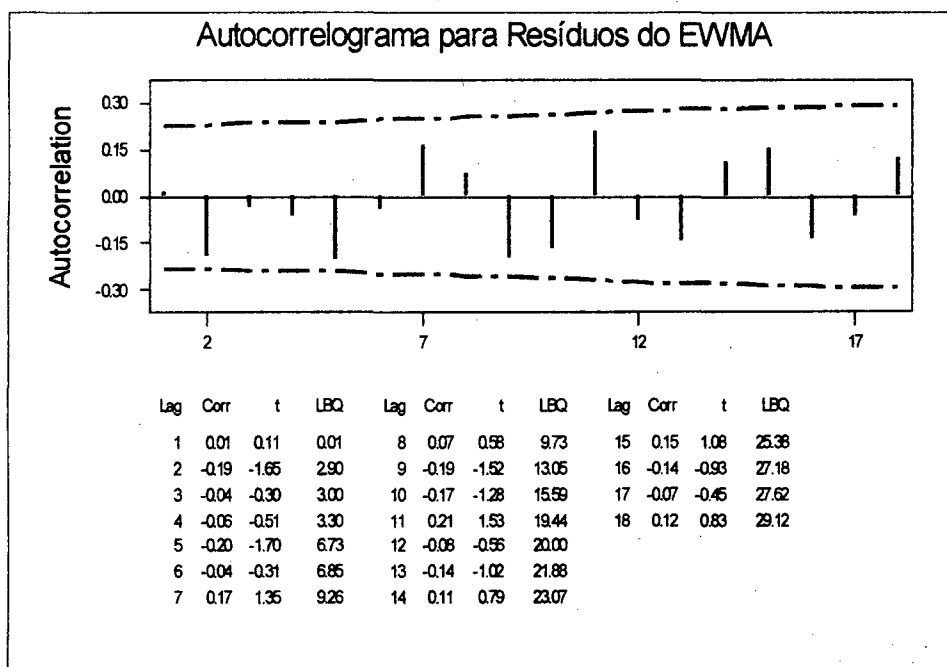


Figura 3-14 – Autocorrelograma para os resíduos do EWMA no caso do peso molecular de um polímero.

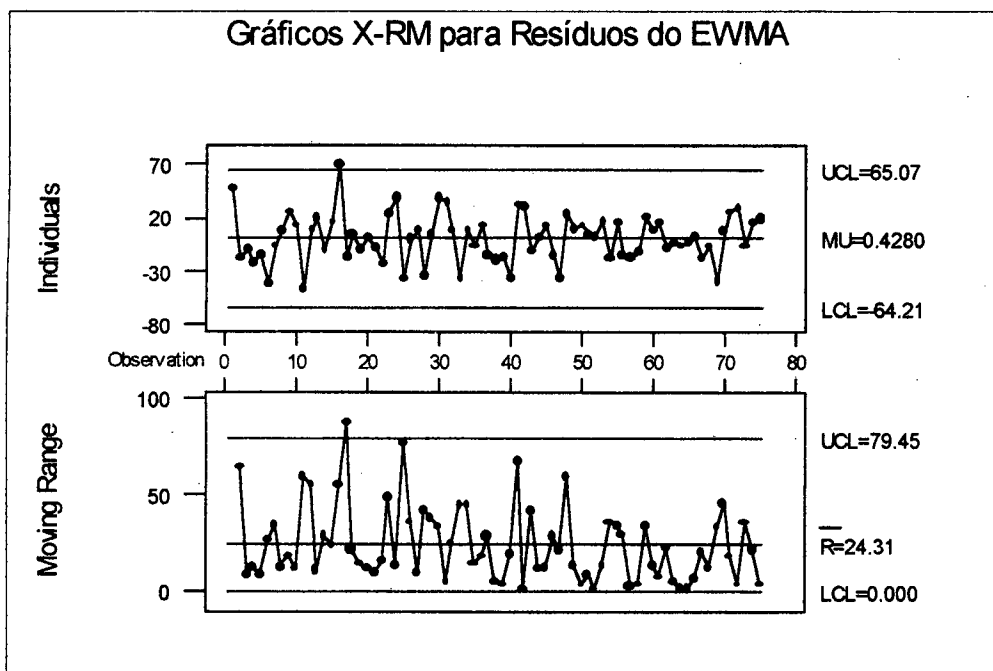


Figura 3-15 – Gráficos X-RM para os resíduos do EWMA no exemplo do peso molecular de um polímero.

3.3-) Solução de Wheeler para Dados Autocorrelacionados

Vamos apresentar neste capítulo mais uma solução para CEP com autocorrelação, sugerida por WHEELER (1994).

3.3.1-) Cálculo dos Limites de Controle

Segundo WHEELER (1994), a autocorrelação tem dois efeitos visíveis sobre os gráficos de controle:

- a) *Efeito sobre o padrão exibido pelos pontos do gráfico:* por inspeção visual, observa-se que os dados exibem padrões de comportamento não aleatório, como ciclos e tendências. É necessário que se tenha um conhecimento profundo do processo, para que seja possível diferenciar quando a autocorrelação está presente ou quando realmente existe a ação de uma causa especial.
- b) *Efeito sobre os limites de controle:* os limites calculados segundo fórmulas tradicionais se tornam mais estreitos na presença de autocorrelação, devendo ser recalculados. Já notamos este ponto em nosso trabalho, na seção 2.3.

Esses dois aspectos somente são significativos, ainda segundo WHEELER (1994), quando a autocorrelação for grande (“excessiva”) e positiva para o lag 1 (r_1 maior ou igual a 0,8). Coeficientes de autocorrelação pequenos teriam um efeito visível pequeno sobre o padrão exibido pelas observações do gráfico e os limites de controle.

Este ponto é corretamente refutado, em nossa opinião, por autores como MARAGAH & WOODALL (1992) e ALWAN (1992). Segundo estes autores, níveis modestos de autocorrelação (como 0,25) introduzem uma grande quantidade de alarmes falsos, deteriorando rapidamente as propriedades estatísticas dos gráficos de controle. Apesar de não se alterar muito o padrão exibido pelo gráfico ou seus limites de controle para coeficientes de autocorrelação pequenos, a performance já fica bastante comprometida nestes casos.

Sem justificar seu procedimento (ou seja, utilizando uma expressão empírica), Wheeler sugere a seguinte correção para a variabilidade do processo para efetuar o ajuste dos limites de controle:

$$\hat{\sigma}_{\text{Wheeler}}(X) \approx \frac{\overline{\text{RM}}}{d_2 \sqrt{1-r_1^2}} \quad (3-19)$$

Onde r_1 é o coeficiente de autocorrelação para o lag 1. O uso da estimativa acima é equivalente a ampliar os limites naturais usuais do processo por um fator de:

$$F = \frac{1}{\sqrt{1-r_1^2}} \quad (3-20)$$

Isto nos fornece as seguintes expressões para os limites de controle de um gráfico de valores individuais (X), o que constitui a chamada **Solução de Wheeler para CEP com autocorrelação**.

$$\begin{aligned} \text{LSC}_{\text{Wheeler}} &= \bar{x} + \frac{3 \overline{\text{RM}}}{d_2 \sqrt{1-r_1^2}} \\ \text{LC}_{\text{Wheeler}} &= \bar{x} \\ \text{LIC}_{\text{Wheeler}} &= \bar{x} - \frac{3 \overline{\text{RM}}}{d_2 \sqrt{1-r_1^2}} \end{aligned} \quad (3-21)$$

Notamos que o efeito do aumento dos limites de controle segundo a expressão acima começa a ser notado com maior intensidade para valores de r_1 iguais ou maiores a 0,5 ($F = 1,154$ aproximadamente, calculado segundo (3-20)).

3.3.2-) Exemplos da Solução de Wheeler

Passamos a analisar dois exemplos que já vimos no capítulo anterior sobre aplicações do EWMA para autocorrelação e aplicar a solução de Wheeler, verificando se seus resultados são compatíveis.

Exemplo de Aplicação 3.5

No exemplo de aplicação 3.3 da seção 3.2.2, sobre a concentração de um produto químico, observamos no gráfico X vários pontos fora dos limites de controle. Para estes dados obtivemos os seguintes valores de parâmetros estatísticos:

$$\overline{RM} = 3,636$$

$$\hat{\sigma}(X) = 3,223$$

Os limites tradicionais do gráfico de X ficam em:

$$LSC = 209,7$$

$$LC = 200,0$$

$$LIC = 190,3$$

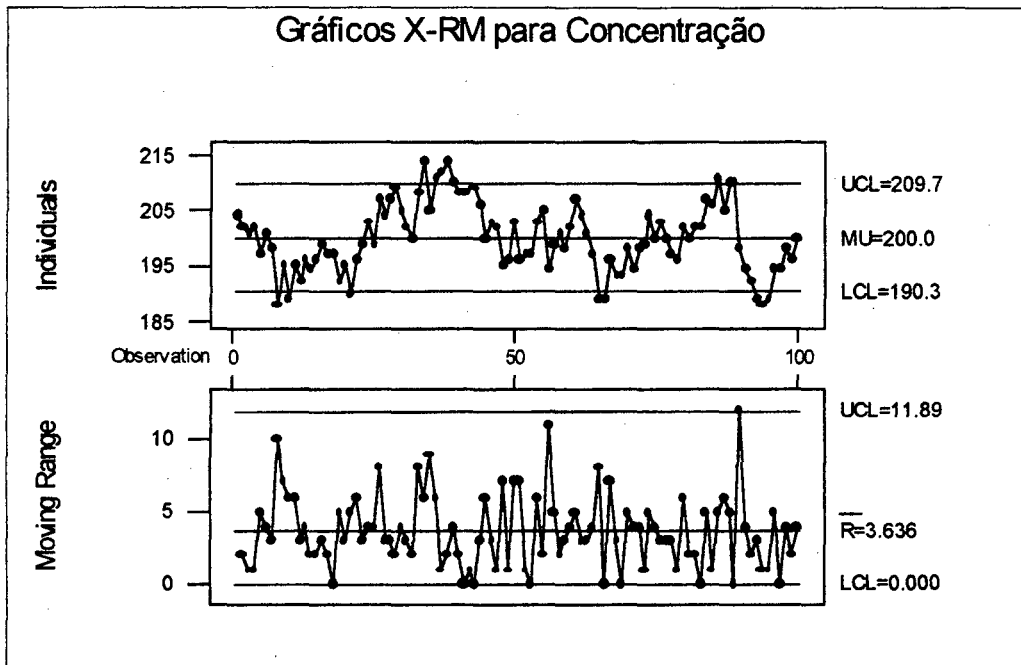


Figura 3-16 – Gráficos X-RM para a concentração com limites convencionais (não ajustados para autocorrelação).

A estimativa do coeficiente de autocorrelação para o lag 1 é $r_1 = 0,75$. Utilizando a fórmula (3-19) para fazer uma nova estimativa de $\sigma(X)$ na presença de autocorrelação e considerando o valor $d_2 = 1,128$ para $n = 2$ (DUNCAN, 1985), temos:

$$\hat{\sigma}_{\text{Wheeler}}(X) = \frac{3,636}{(1,128)\sqrt{1-(0,75)^2}} = 4,873$$

Portanto, os novos limites do gráfico de valores individuais (figura 3-15), calculados pelas fórmulas (3-20), passarão a ser:

$$LSC_{\text{Wheeler}} = 214,6$$

$$LC_{\text{Wheeler}} = 200,0$$

$$LIC_{\text{Wheeler}} = 185,4$$

Verificamos que o gráfico obtido na figura (3-17) mostra uma situação de controle estatístico. A única ressalva são as observações 34 e 38, que estão próximas ao limite superior de controle. Estas observações podem ser tomadas como uma advertência e não como indicadores de causa especial de variação. O resultado obtido com a solução de Wheeler coincide, portanto, com aquele obtido por nós anteriormente com outras técnicas.

Observamos também que na solução de Wheeler não é possível utilizar regras de sensibilidade na análise do gráfico pois os dados são autocorrelacionados.

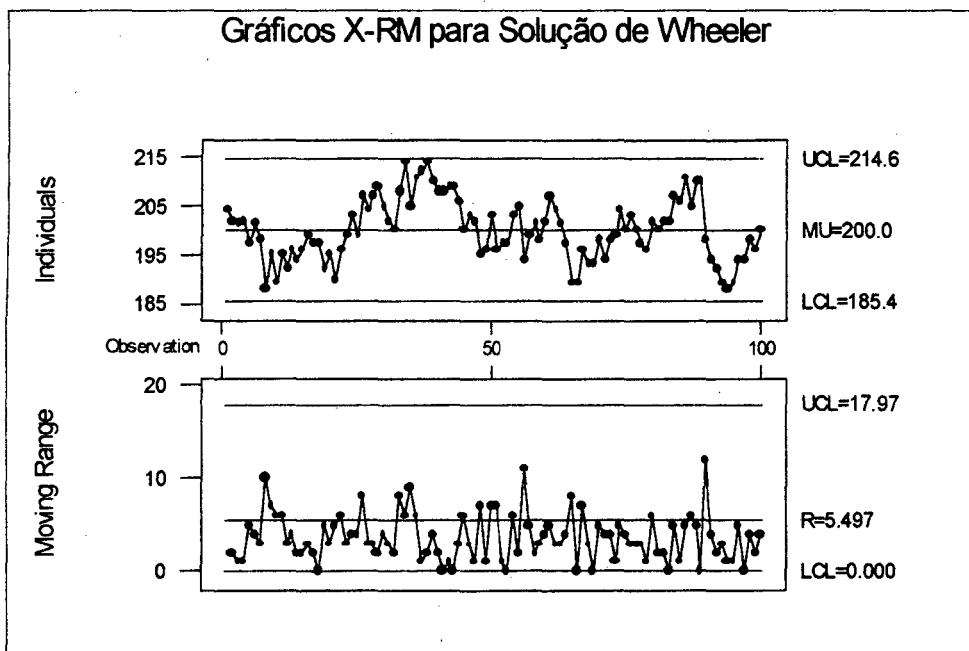


Figura 3-17 – Gráficos X-RM com limites ajustados para autocorrelação segundo a solução de Wheeler.

Exemplo de aplicação 3.6

Vamos analisar novamente o exemplo de aplicação 3.1 do peso molecular de um polímero. A figura (3-18) apresenta os gráficos X-RM convencionais.

As propriedades estatísticas para este processo são as seguintes:

$$\overline{RM} = 17,81$$

$$\hat{\sigma}(X) = 15,78$$

Gráfico de X:

$$LSC = 2048$$

$$LC = 2001$$

$$LIC = 1954$$

A estimativa para o coeficiente de autocorrelação r_1 é 0,66.

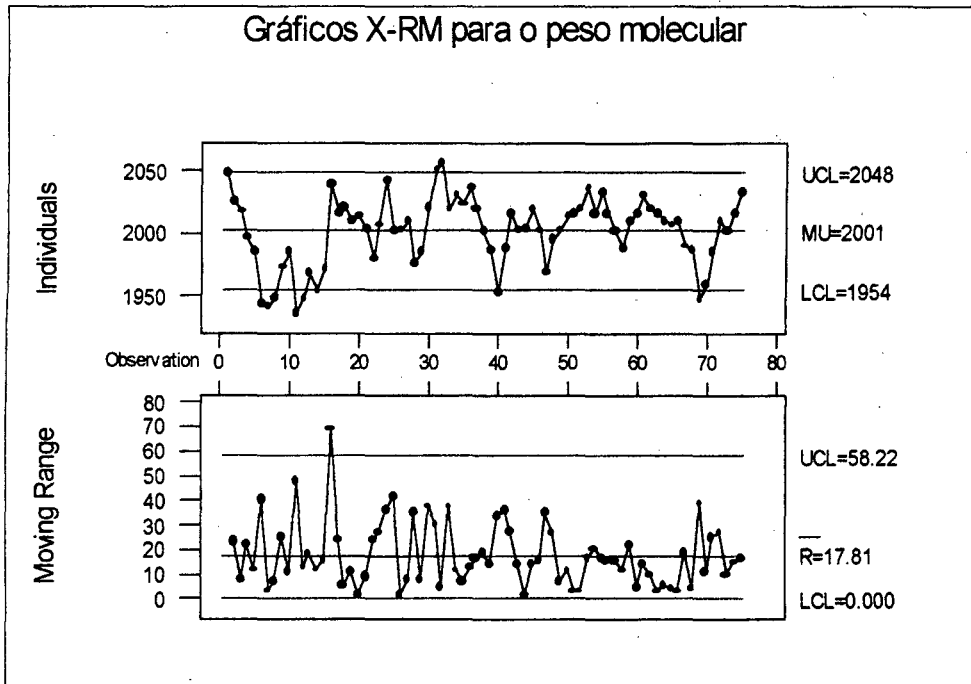


Figura 3-18 – Gráficos X-RM com limites naturais não ajustados para a autocorrelação ($r_1 = 0,66$).

Ajustando a estimativa de $\sigma(X)$ para o processo com autocorrelação, podemos calcular os novos limites de controle e traçar um novo gráfico.

$$\hat{\sigma}_{Wheeler}(X) = \frac{17,81}{(1,128) \sqrt{1 - (0,66)^2}} = 21,01$$

Limites de controle para o gráfico de X:

$$LSC_{Wheeler} = 2064$$

$$LC_{Wheeler} = 2001$$

$$LIC_{Wheeler} = 1938$$

Observamos que no gráfico X da solução de Wheeler, mostrados na figura (3-19), há um ponto abaixo do limite inferior de controle, correspondente à observação 12, o que indica que deve haver causa especial. Com as soluções EWMA e AR(1) obtivemos também um pontos fora dos limites, porém o mesmo correspondia à observação 16 e era um ponto acima do limite superior.

Portanto, consideramos que a solução de Wheeler não é equivalente às demais, embora não estejamos em condições de avaliar as diferenças de performance entre elas. Tentaremos fazer isto no próximo capítulo.

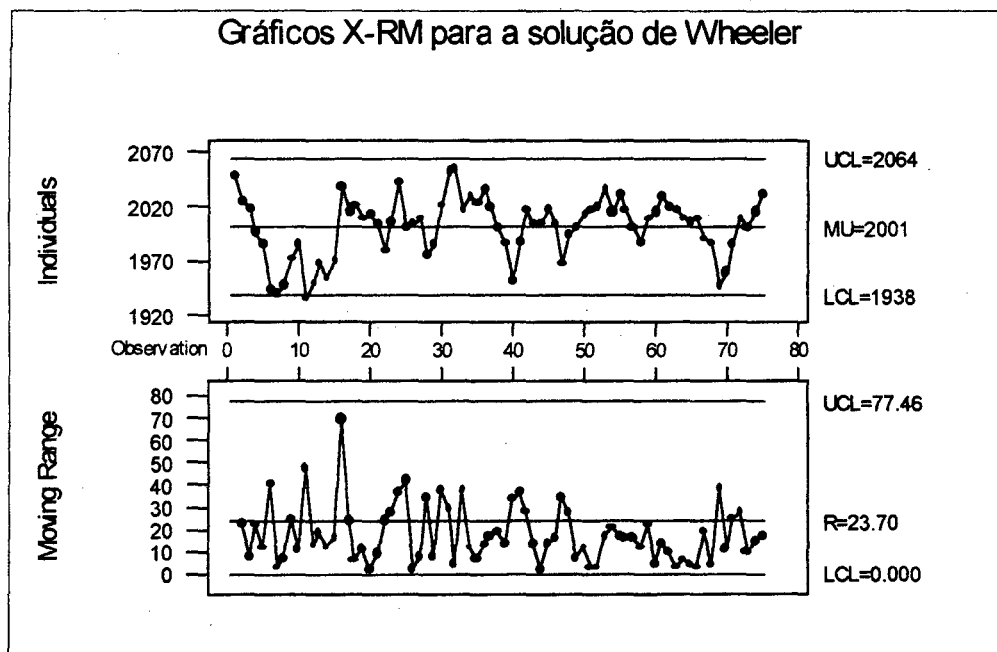


Figura 3-19 – Gráficos X-RM para o peso molecular com limites ajustados segundo a solução de Wheeler.

Capítulo 4 – Análise Comparativa da Performance dos Gráficos de Controle com Dados Autocorrelacionados

4.1-) Especificação do Teste

Nesta parte fazemos, através de simulação computacional, uma análise comparativa da performance (em termos da ARL sob controle e fora de controle) dos gráficos de controle estudados para dados autocorrelacionados, dando ao final recomendações sobre o seu uso. A simulação foi desenvolvida em Microsoft Excel 2000, utilizando a linguagem de macros Visual Basic for Applications (VBA) para os cálculos e modelos de planilha (“templates”) para armazenar os resultados. O código fonte, bem como os modelos de planilhas estão a disposição dos interessados em reproduzir a simulação. As soluções analisadas foram:

- a) Gráfico de Shewhart convencional para valores individuais (ver seção 1.3.2), utilizando os dados originais, autocorrelacionados.
- b) Gráfico de Shewhart para valores individuais dos resíduos da EWMA (ver seção 3.2.2).
- c) Gráfico de Shewhart para valores individuais dos resíduos para um modelo de séries temporais (ver seção 3.1.2).
- d) Gráfico de Shewhart para valores individuais, com limites modificados segundo a solução de Wheeler (ver seção 3.3.1).

As soluções foram testadas para séries temporais estacionárias com dados autocorrelacionados. Os modelos de séries temporais que aparecem mais frequentemente na prática, segundo MONTGOMERY & RUNGER (1994), são o AR(1), AR(2) (vistos no capítulo 3 deste trabalho), além de MA(1), MA(2) (modelos com médias móveis de 1ª e 2ª ordem, respectivamente) e ARMA(1,1) (modelo autorregressivo de 1ª ordem com média móvel de 1ª ordem). Estes foram os modelos que utilizamos para gerar as séries da simulação.

Uma descrição das propriedades dos modelos que não vimos no capítulo 3 é feita por diversos autores (entre eles VANDAELE(1983) e MAKRIDAKIS(1983)) e os parâmetros dos modelos MA(1), MA(2) e ARMA(1,1) e fórmulas do desvio padrão são dados a seguir.

- MODELO MA(1)

Expressão de x_t : $x_t = \mu + a_t - \theta_1 a_{t-1}$ (4-1)

Sendo θ_1 um parâmetro do processo

$$\sigma(x) = \sigma_a \sqrt{1 + \theta_1^2}$$

- MODELO MA(2)

Expressão de x_t : $x_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ (4-2)

Sendo θ_1 e θ_2 parâmetros do processo.

$$\sigma(x) = \sigma_a \sqrt{1 + \theta_1^2 + \theta_2^2}$$

- MODELO ARMA(1,1)

Expressão de x_t : $x_t = (1 - \phi_1) \mu + \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1}$ (4-3)

Sendo ϕ_1 e θ_1 parâmetros do processo.

$$\sigma(x) = \sigma_a \sqrt{\frac{1 - 2\phi_1\theta_1 + \theta_1^2}{1 - \phi_1^2}}$$

A seguir descrevemos as etapas do teste de simulação realizado.

Etapas do teste de simulação

1-) Geramos, através de algoritmo computacional, uma seqüência de dados aleatória que segue um dado modelo de séries temporais. Criamos uma variável aleatória normal a_t com média 0 e desvio padrão 1, utilizando-se para tanto as funções internas do Microsoft Excel. O modelo de séries temporais foi gerado a partir da série a_t , com as combinações de parâmetros dadas na tabela (4-1). As combinações particulares para a série AR(2) foram escolhidas porque induziram um significativo grau de autocorrelação no processo.

Combinação de Parâmetros	Modelo AR(1)	Modelo AR(2)	Modelo MA(1)	Modelo MA(2)	Modelo ARMA(1,1)
1	$\phi_1 = 0,1$	$\phi_1 = 0,1$ $\phi_2 = 0,8$	$\theta_1 = -0,3$	$\theta_1 = -0,3$ $\theta_2 = -0,1$	$\phi_1 = 0,1$ $\theta_1 = -0,1$
2	$\phi_1 = 0,3$	$\phi_1 = 0,3$ $\phi_2 = 0,6$	$\theta_1 = -0,5$	$\theta_1 = -0,3$ $\theta_2 = -0,9$	$\phi_1 = 0,1$ $\theta_1 = -0,9$
3	$\phi_1 = 0,5$	$\phi_1 = 0,5$ $\phi_2 = 0,4$	$\theta_1 = -0,7$	$\theta_1 = -0,7$ $\theta_2 = -0,1$	$\phi_1 = 0,5$ $\theta_1 = -0,1$
4	$\phi_1 = 0,7$	$\phi_1 = 0,7$ $\phi_2 = 0,2$	$\theta_1 = -0,9$	$\theta_1 = -0,7$ $\theta_2 = -0,9$	$\phi_1 = 0,5$ $\theta_1 = -0,9$
5	$\phi_1 = 0,9$			$\theta_1 = -0,9$ $\theta_2 = -0,9$	$\phi_1 = 0,9$ $\theta_1 = -0,1$
6					$\phi_1 = 0,9$ $\theta_1 = -0,9$

Tabela 4-1 – Combinações de parâmetros para os modelos de séries temporais utilizados na simulação.

2-) Foi induzido um desvio sustentado (degrau) na média do processo após 100 observações da série temporal (de um total de 500 observações), levando o processo para fora de controle estatístico.

3-) Verificamos o número médio de observações após o desvio da média acima até que cada um dos gráficos estudados sinalizasse um ponto fora dos limites de controle (detecção do desvio). O número máximo foi fixado pelo número de pontos gerados em cada série; por ex.: se nenhum ponto esteve fora dos limites até a observação 500, então o valor retornado foi 400.

4-) Repetimos 100 vezes os passos 1, 2 e 3 de nosso experimento para cada combinação de parâmetros e desvios na média de cada série temporal, obtendo o valor e um intervalo de confiança (95% aproximadamente) para a média dos 100 experimentos (ARL). Foi também testada a taxa de alarmes falsos, que dá a performance dos gráficos sob controle, não se executando o passo 2 da seqüência acima.

Os valores das estimativas de ARL obtidas nas simulações realizadas encontram-se no apêndice A. Na seção 4.2 fazemos uma síntese dos resultados obtidos para facilitar a análise e comparação dos métodos.

4.2-) *Análise dos Resultados Obtidos*

Vamos analisar os resultados obtidos para a ARL em três situações: sob controle estatístico (ARL_0) e fora de controle (ARL_1) com desvios na média de $1,5\sigma$ ($k = 1,5$) e de 3σ ($k = 3$).

4.2.1-) ARL_0

Esta seção referencia os gráficos da figura (4-1) comparativos entre as quatro soluções para a ARL sob controle.

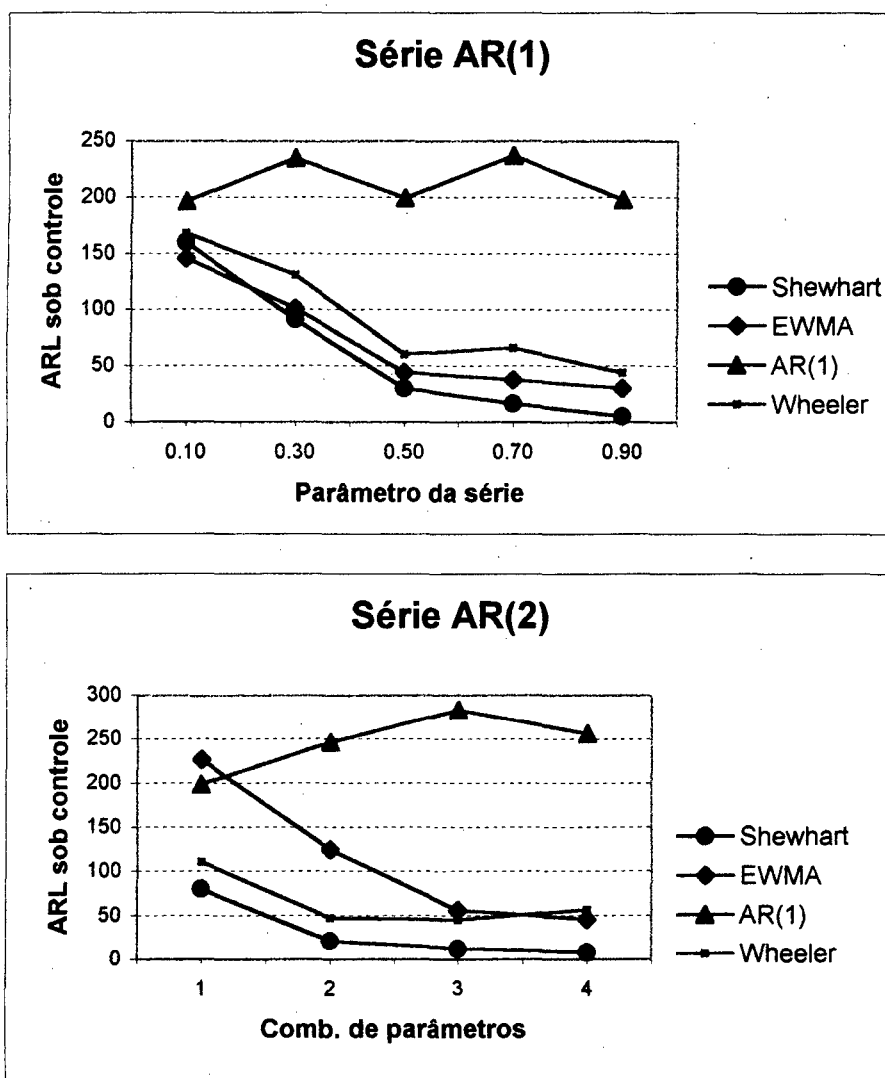


Figura 4-1 – Gráficos comparativos da ARL_0 para as 4 soluções e várias séries geradas, em função da combinação de parâmetros utilizada.

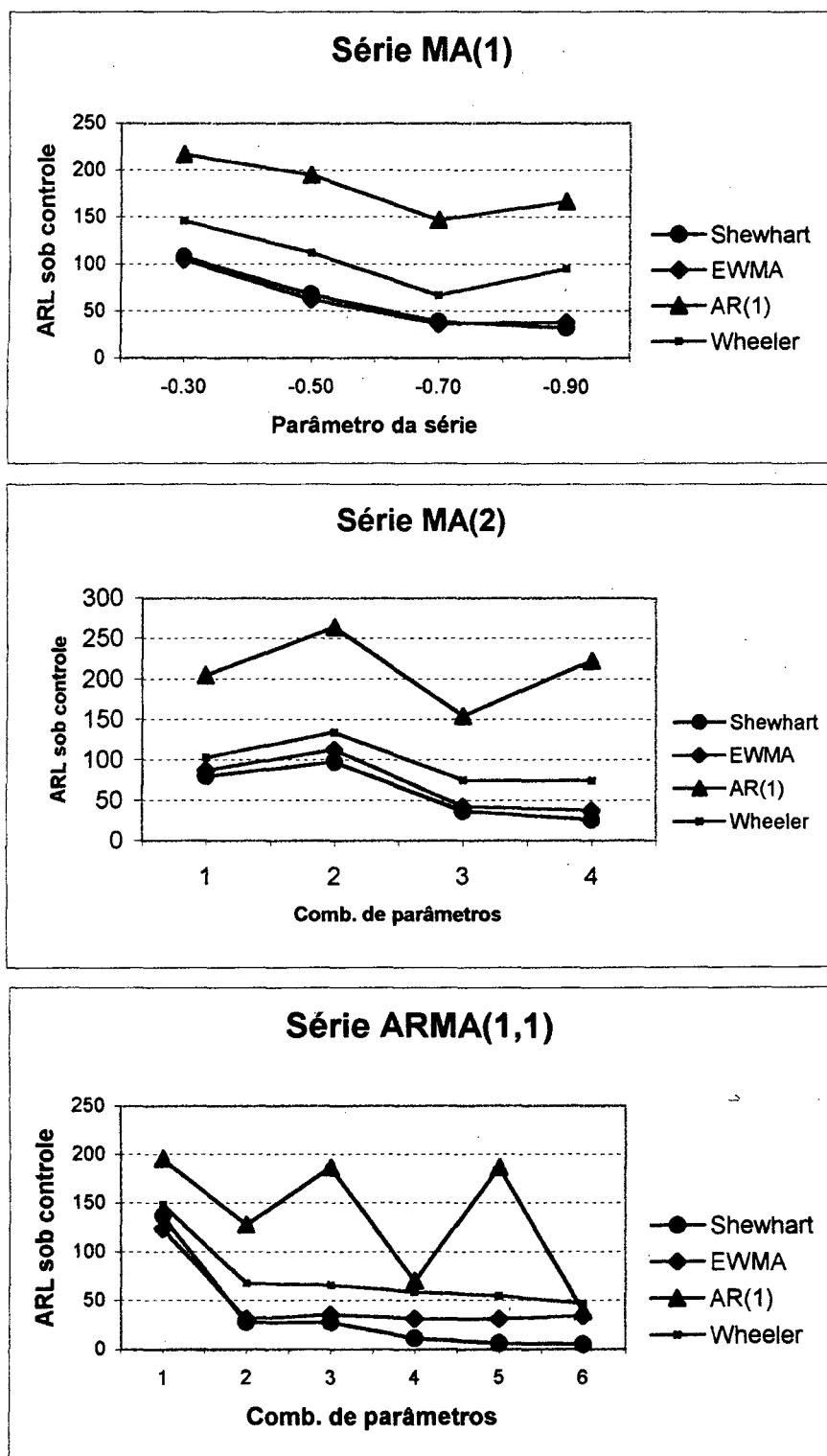


Figura 4-1 (cont.) – Gráficos comparativos da ARL_0 para as 4 soluções e várias séries geradas.

Observamos que a ARL_0 obtida por análise do gráfico de Shewhart é a menor entre os quatro métodos, independente da série analisada. Ela inclusive tende a ter valores menores para valores maiores dos parâmetros das séries, que estão relacionados a uma maior dependência em relação a observações anteriores. Este decréscimo da ARL para valores maiores dos parâmetros das séries na verdade ocorre também com os gráficos de resíduos do EWMA e da solução de Wheeler.

A solução EWMA apresentou em geral uma ARL_0 menor que a solução de Wheeler, porém estando sempre acima daquela do gráfico de Shewhart. No entanto, para determinadas combinações de parâmetros (por exemplo $\phi_1 = 0,1$ e $\phi_2 = 0,8$ da série AR(2)), a ARL_0 da EWMA assume valores maiores (neste caso é 227).

O método do AR(1) é o que apresenta a maior ARL_0 na maioria das situações. Apenas nos casos do MA(2) com parâmetro $\theta_1 = -0,7$ e $\theta_2 = -0,1$ e ARMA(1,1) nas combinações em que $\theta_1 = -0,9$ a sua ARL_0 cai.

Portanto, concluímos por este estudo que o AR(1) é o método com menor taxas de alarmes falsos, seguido pelo método de Wheeler.

4.2.2-) ARL_1 para $k=1,5$

Para a ARL fora de controle estatístico com desvio em relação à média de $1,5\sigma$, estamos nos referindo aos gráficos da figura (4-2).

Na seção anterior, concluímos que o gráfico de Shewhart é uma solução inviável devido a uma taxa muito alta de alarmes falsos. Portanto, nesta e na próxima seção, embora o gráfico de Shewhart faça parte de nossa análise, não o estaremos recomendando mesmo que tenha um excelente desempenho para detectar desvios na média do processo.

O gráfico de Wheeler tem um desempenho similar (com uma ARL_1 muito baixa) ao de Shewhart para a maioria das séries utilizadas, exceto AR(2) e ARMA(1,1) nos quais o gráfico de Shewhart leva vantagem. No método de Wheeler a ARL_1 não varia muito com a mudança nos parâmetros das séries.

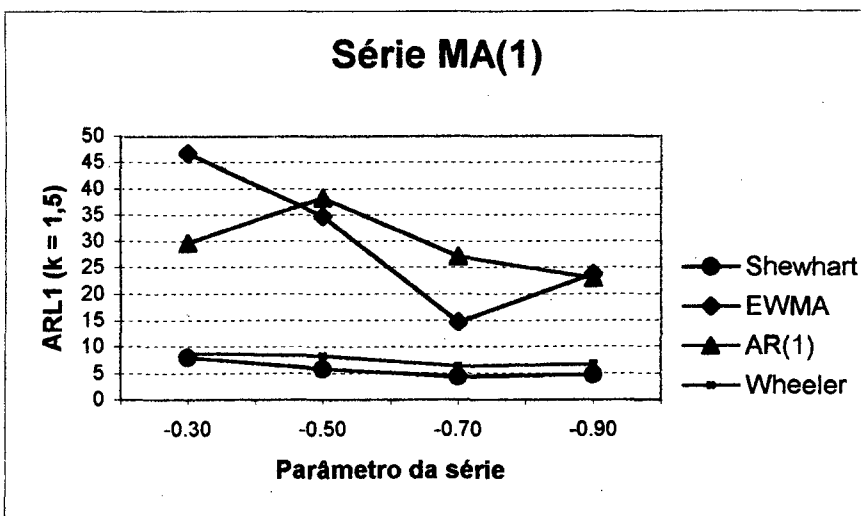
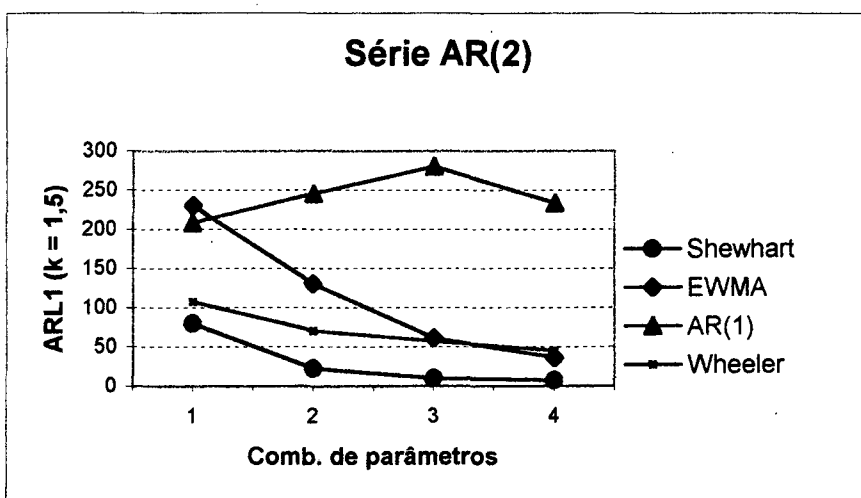
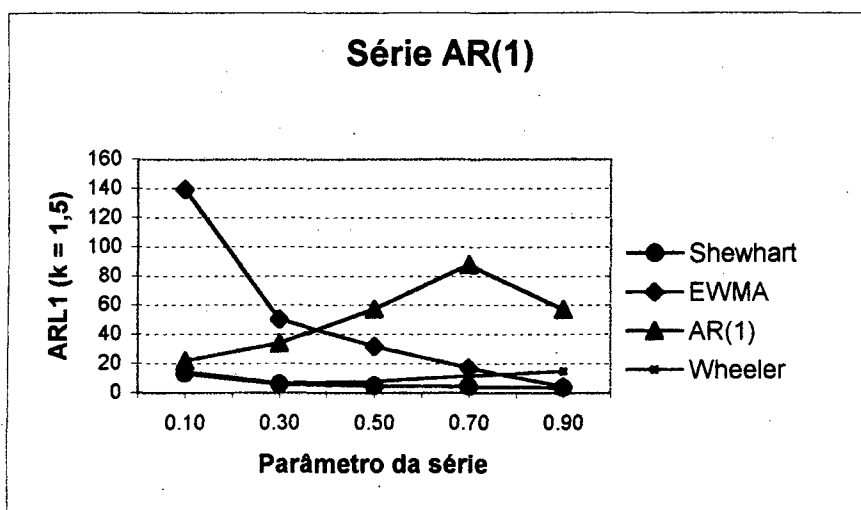


Figura 4-2 – Gráficos comparativos da $ARL_1 (k = 1,5)$ para as 4 soluções e séries geradas.

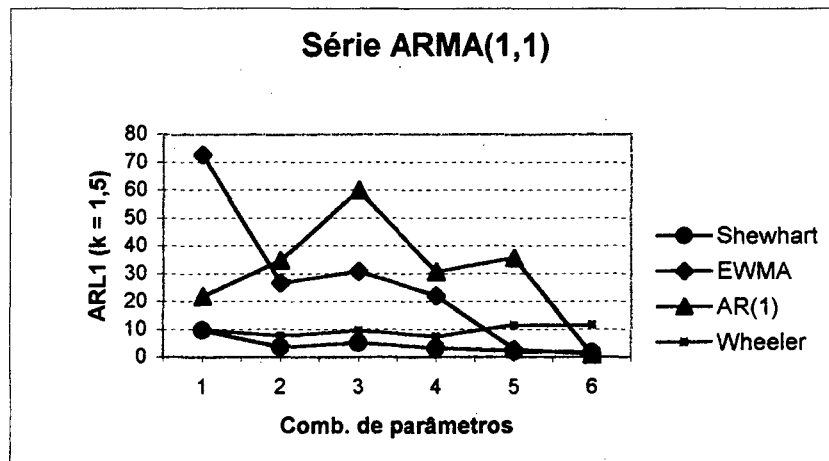
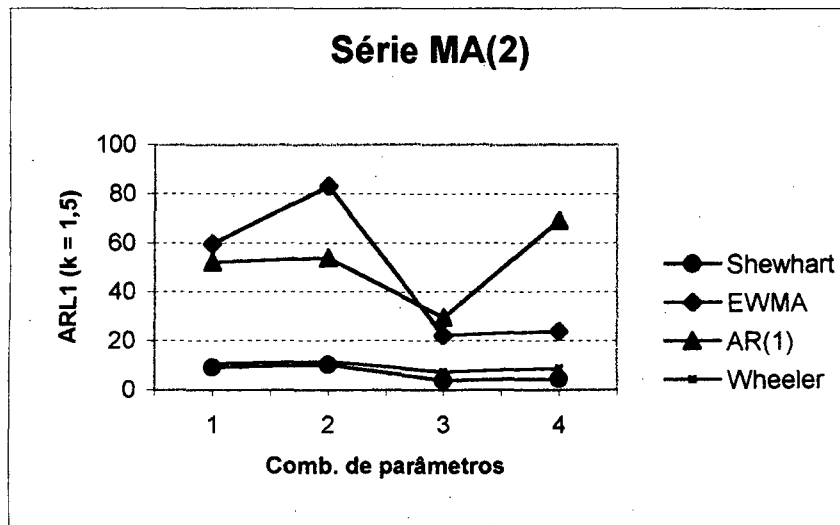


Figura 4-2 (cont.) – Gráficos comparativos da ARL_1 ($k = 1,5$) para as 4 soluções e séries geradas.

Considerando os gráficos de resíduos da EWMA e AR(1), notamos que há uma alternância entre eles: ora a solução AR(1) tem um desempenho melhor, ora a EWMA é que tem a melhor performance. As situações em que a EWMA leva vantagem são aquelas em que os parâmetros de média móvel (θ) são maiores em módulo. Por exemplo, para a série MA(1) e parâmetro $\theta_1 = -0,7$, a solução baseada na EWMA detecta o desvio na média em 15 observações, enquanto que a solução baseada no AR(1) detecta o desvio em 27 observações. Já para ARMA(1,1) com parâmetros $\phi_1 = 0,9$ e $\theta_1 = -0,1$, EWMA detecta em 3 observações contra 36 de AR(1). Isto nos sugere que a EWMA é adequada para séries que seguem modelos de médias móveis.

Portanto, para um desvio na média de $1,5\sigma$, concluímos que a solução de Wheeler é a que apresenta condições mais favoráveis, seguida pelo EWMA. Devemos ressaltar que em nosso estudo não utilizamos as regras de sensibilidade, que poderiam melhorar bastante a performance das soluções baseadas em EWMA e AR(1), já que não é possível aplicá-las à solução de Wheeler porque os dados nesse gráfico são autocorrelacionados.

4.2.3-) ARL_1 para $k=3$

Para desvios na média de 3σ , estamos nos referindo aos gráficos da figura (4-3). Valem as mesmas considerações do item anterior com relação à performance do gráfico de Shewhart, que é inviabilizado pela baixa ARL sob controle.

À exceção do caso de AR(2), para o qual parece que nenhum gráfico exceto o de Shewhart detecta com eficiência um desvio na média de 3σ , a performance das soluções Shewhart e Wheeler é semelhante e os desvios foram detectados após 1 ou 2 observações.

A performance dos gráficos de resíduos da EWMA e AR(1) é similar na maioria dos casos, seguindo o gráfico de Wheeler. A EWMA, principalmente quando os parâmetros das séries temporais assumem valores baixos, tem sua performance diminuída, como no caso do AR(1) com $\phi_1 = 0,1$ ($ARL_1 = 19$) ou no ARMA(1,1) com $\phi_1 = 0,1$ e $\theta_1 = -0,1$ ($ARL_1 = 15$). Isto nos sugere que a solução EWMA é melhor para desvios de até 2σ do que para desvios grandes (maiores que 2σ).

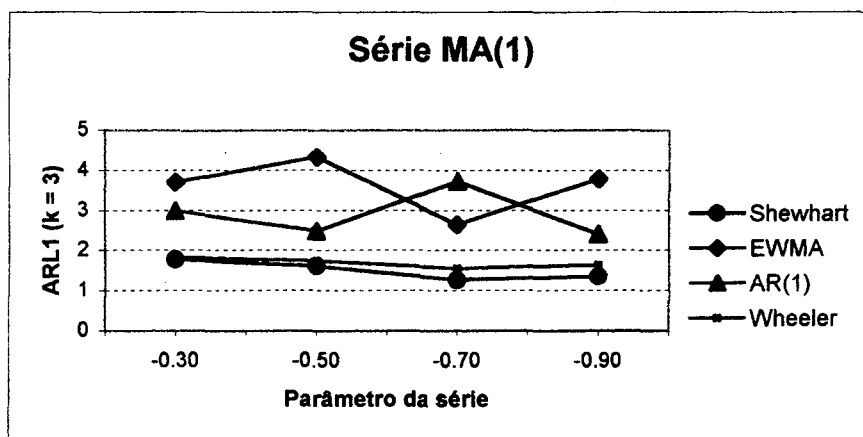
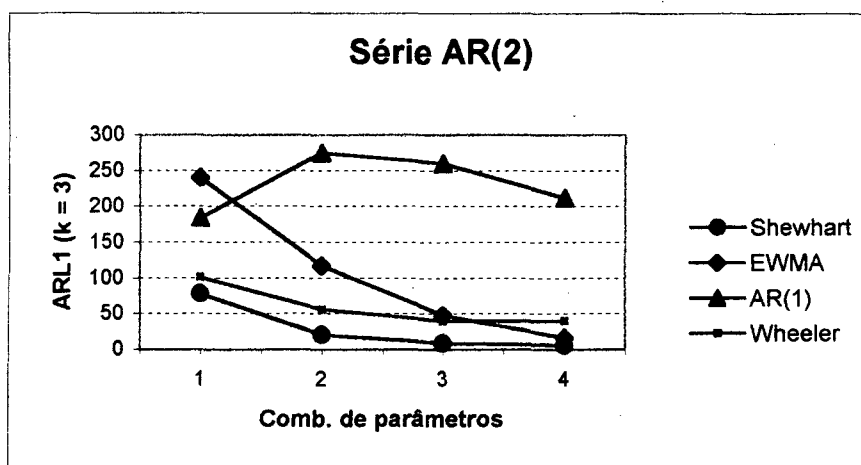
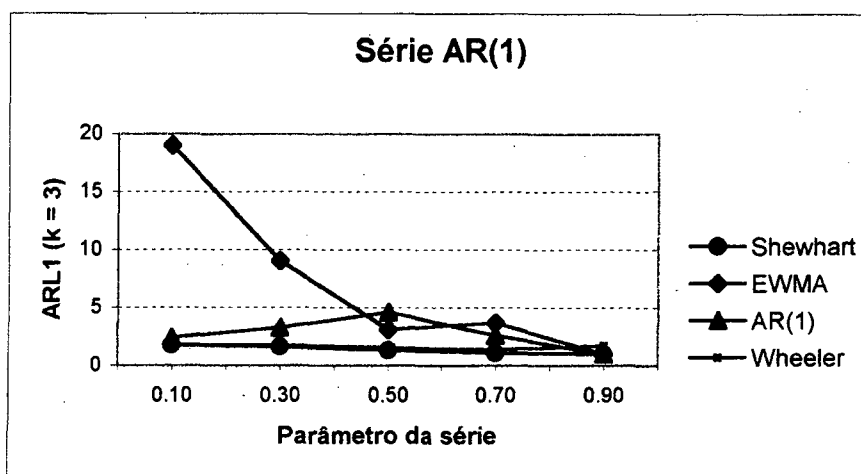


Figura 4-3 – Gráficos comparativos da $ARL_1 (k = 3)$ para as 4 soluções e séries geradas.

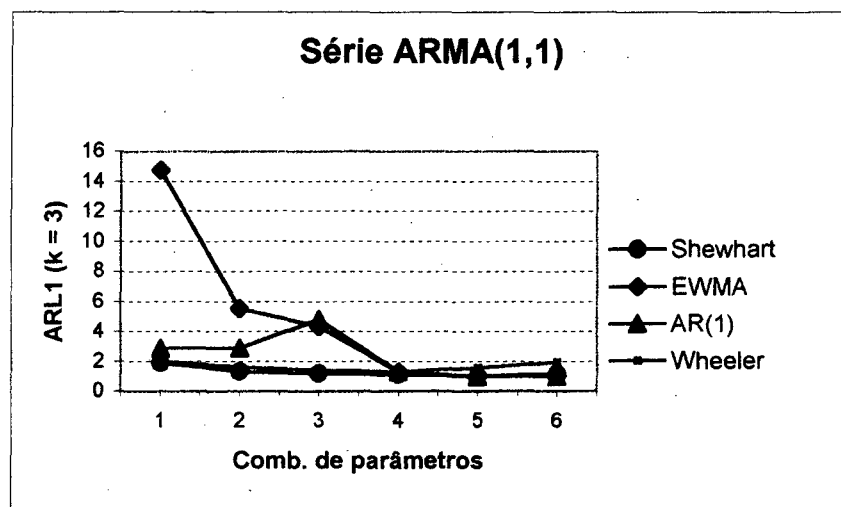
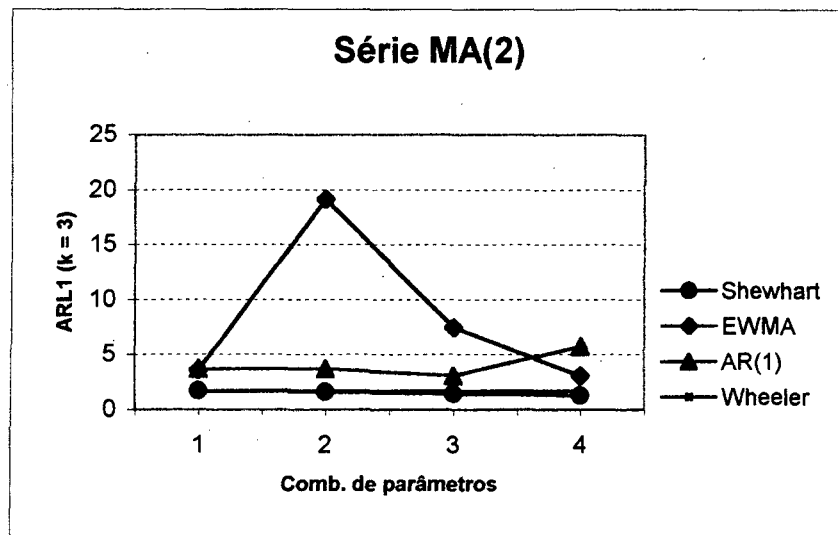


Figura 4-3 (cont.) – Gráficos comparativos da ARL_1 ($k = 3$) para as 4 soluções e séries geradas.

4.2.4-) Conclusão

Apresentamos a seguir nossa conclusão a respeito deste estudo e de nosso trabalho.

O gráfico de Shewhart não deve ser utilizado em presença de autocorrelação, porque geralmente irá apresentar uma taxa de alarmes falsos inaceitável.

A solução de Wheeler (expansão dos limites de controle) é uma alternativa viável ao gráfico de Shewhart, tendo inclusive uma ARL_0 vantajosa. Sua maior desvantagem é a de apresentar os dados autocorrelacionados e não permitir o uso de regras de sensibilidade.

A solução EWMA não apresentou em nosso estudo uma ARL_0 muito vantajosa. No entanto, apresentou um bom desempenho na detecção de pequenos desvios da média, mesmo na presença de autocorrelação. Esta performance pode ser ainda melhor com o uso de regras de sensibilidade.

A melhor solução, de acordo com a nossa avaliação, é a baseada nos gráficos de resíduos de AR(1), pois é a que apresenta maior ARL sob controle e tem muito bom desempenho na detecção de desvios grandes da média, embora não seja tão bom nos pequenos desvios, resposta esta que pode melhorar com o uso de regras de sensibilidade.

Com relação à facilidade de implementação dos métodos estudados na indústria, podemos considerar o de Wheeler e EWMA como mais simples pois não envolvem a determinação de um modelo de séries temporais para os dados. Esta é justamente a dificuldade de se implementar o AR(k); porém com os recursos computacionais atuais esta dificuldade tende a ser minimizada, viabilizando seu uso efetivo.

Gostaríamos de sugerir outras abordagens para complementar este trabalho, entre elas:

- o estudo destes métodos que utilizasse regras de sensibilidade para uma melhor comparação, entre as diversas alternativas;
- o uso de modelos de séries temporais mais genéricos que o AR(k), como os da classe ARIMA, para maior precisão na determinação do modelo;
- o estudo de outros tipos de desvio, como mudanças na variabilidade do processo;
- o tratamento da autocorrelação utilizando outras soluções, como diferentes estimadores para a variabilidade do processo (S/c_4) e o uso de controle automático de processos.

Bibliografia

MONTGOMERY, D. C. - *Introduction to Statistical Quality Control*, 3rd Edition, New York, Wiley, 1996.

WHEELER, D. J. - *Advanced Topics in Statistical Process Control*, Knoxville, SPC Press, 1994.

DUNCAN, A. J. - *Quality Control and Industrial Statistics*, 5th Edition, Homewood, Richard. D. Erwin, 1985.

SHEWHART, W. A. - *Economic Control of Quality of Manufactured Product*, New York, Van Nostrand, 1931.

RAMOS, A. W. - *Uma Contribuição aos Estudos de Capacidade de Máquina*, Tese de Doutorado, São Paulo, 1999.

SCHNEIDER, H. & PRUETT, J. M. - *Essentials of SPC in the Process Industries*, Research Triangle Park, Instrument Society of America, 1993.

SHAW, J. A. - "Statistical Process Control for the Process Industries", *ISA Transactions*, 30 (1), p. 99-106, 1991.

PYSDEK, T. - *Pysdek's Guide to SPC, Milwaukee*, ASQC Quality, 1992.

NEWBOLD, P. - *Statistics for Business and Economics*, 4th Edition, Englewood Cliffs, Prentice Hall, 1995.

CRYER, J. D. & RYAN, T. P. - "The estimation of Sigma for an X Chart: \overline{MR}/d_2 ou S/c_4 ?", *Journal of Quality Technology*, 22 (3), p. 187-191, 1990.

VASILOPOULOS, A. V. & STAMBOULIS, A. P. - "Modification of Control Chart Limits in the Presence of Data Correlation", *Journal of Quality Technology*, 10 (1), p. 20-30, 1978.

ALWAN, L. C. & ROBERTS, H. V. – “Time-Series Modeling for Statistical Process Control”, *Journal of Business and Economic Statistics*, 6 (1), p. 87-95, 1988.

MONTGOMERY, D. C. & MASTRANGELO, C. M. – “Some Statistical Process Control Methods for Autocorrelated Data”, *Journal of Quality Technology*, 23 (3), p. 179-192, 1991.

MARAGAH, H. D. & WOODHALL, W. H. – “The Effect of Autocorrelation on the Restrospective X-Chart”, *Journal of Statistical Computation and Simulation*, 40, p. 29-42, 1992.

VANDAELE, W. – *Applied Time Series and Box-Jenkins Models*, Orlando, Academic Press, Inc. 1983

MAKRIDAKIS, S. G. – *Forecasting : Methods and Applications*, 2nd Edition, New York, Wiley, 1983.

MONTGOMERY, D. C. & RUNGER, G. C. – *Statistical Process Control for the Chemical and Process Industries*, Kirkland, Statistical Productivity, 1994

ROBERTS, S. W. – “Control Chart Tests Based on Geometric Moving Averages”, *Technometrics*, 1 (3), p. 239-251, 1959.

HUNTER, J. S. – “The Exponentially Weighted Moving Average”, *Journal of Quality Technology*, 21 (3), p.155-162, 1986.

CROWDER, S. V. - “Design of Exponentially Weighted Moving Average, *Journal of Quality Technology*, 21 (3), p. 155-162. 1989.

LUCAS, J. M. & SACCUCCI, M. S. – “Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements”, *Technometrics*, 32 (1), p. 1-29, 1990.

ALWAN, L. C. – “Effects of Autocorrelation on Control Chart Performance”, *Commun. Statistics – Theory and Methods*, 21 (4), p. 1025-1049, 1992.

SCHNEIDER, H. & PRUETT, J. M. – “Control Charting Issues in the Process Industries”, *Quality Engineering*, 6 (3), p. 347-373, 1994.

RIGDON, S. E. et. al. – “Design Strategies for Individual and Moving Range Control Charts”, *Journal of Quality Technology*, 26 (4), p. 274-287, 1994.

BERTHOUEX, P. M. et. al. – “Monitoring Sewage Treatment Plants: Quality Control Aspects”, *Journal of Quality Technology*, 10 (4), p. 139-149, 1978.

WHEELER, D. J. – “Detecting a Shift in Process Average: Tables of the Power Function for X bar Charts”, *Journal of Quality Technology*, 15 (4), p. 155-170, 1983.

RAMOS, A. W. – “CEP para Processos Contínuos e em Bateladas”, *Estudos de Caso da Apostila de Treinamento da Optima Engenharia e Consultoria S/C Ltda.*, p. 16 e seguintes.

CROWDER, S. V. – “A Simple Method for Studying Run-Lenght Distributions of Exponentially Weighted Moving Average Charts”, *Technometrics*, 29 (4), p. 401-407, 1987.

ALBIN, L. A. et. al. – “An X and EWMA Chart for Individual Observations”, *Journal of Quality Technology*, 29 (1), p. 41-48, 1997.

NG, C. H. & CASE, K. E. – “Development and Evaluation of Control Charts Using Exponentially Weighted Moving Average”, *Journal of Quality Technology*, 21 (4), p. 242-250, 1989.

***Apêndice A – Tabelas Geradas pela Simulação
da Performance dos Gráficos de Controle para
Dados Autocorrelacionados***

**Simulação do Modelo AR(1)
Resultados para:**

Phi1 0.1

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	160	146	196	168
0.5	108	160	137	116
1	42	161	70	42
1.5	13	140	22	15
2	6	86	9	6
3	2	19	2	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
130.760	189.920	117.119	175.081	165.272	227.688	138.771	198.169
86.619	128.461	131.624	188.376	113.098	161.662	92.769	138.571
33.530	50.710	130.598	191.242	53.856	86.944	33.879	51.021
10.161	16.779	108.545	170.655	15.836	28.384	10.881	18.239
4.327	6.813	59.395	111.605	6.830	11.770	4.451	6.949
1.464	2.076	4.666	33.414	1.808	3.032	1.494	2.106

Simulação do Modelo AR(1)

Resultados para:

Phi1 0.3

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	92	102	235	132
0.5	60	83	173	78
1	18	73	88	21
1.5	6	51	34	7
2	4	37	14	4
3	2	9	3	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
72.081	111.319	81.403	122.057	205.559	264.161	107.275	155.945
45.633	74.207	63.313	101.867	143.884	202.276	59.823	96.237
13.814	21.486	53.089	93.731	67.080	108.760	16.274	25.506
5.122	7.238	32.496	68.764	23.542	44.838	5.713	7.967
3.179	4.541	19.782	53.538	9.895	18.865	3.498	5.302
1.416	1.924	1.658	16.502	2.113	4.467	1.513	2.087

Simulação do Modelo AR(1)
Resultados para:

Phi1 0.5

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	30	44	199	60
0.5	24	40	174	46
1	14	40	110	22
1.5	5	32	58	8
2	3	21	20	4
3	1	3	5	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
23.545	37.275	30.952	57.768	169.360	229.160	44.972	75.748
19.111	29.789	30.241	49.079	146.263	202.457	36.610	55.330
9.881	17.159	31.458	48.882	87.558	132.722	15.703	28.177
4.132	6.388	22.856	41.344	43.943	71.317	6.585	9.815
2.301	3.419	13.674	27.926	13.998	25.502	3.218	4.922
1.119	1.541	1.549	4.671	2.959	6.301	1.295	1.865

Simulação do Modelo AR(1)
Resultados para:

Phi1 0.7

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	17	38	237	66
0.5	12	30	157	30
1	8	34	131	18
1.5	4	17	88	12
2	2	11	35	5
3	1	4	3	1

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
12.678	20.622	27.533	47.647	206.072	268.028	51.627	81.133
9.086	14.234	22.938	36.422	128.742	185.858	24.126	36.634
6.111	9.509	23.155	43.985	105.016	156.964	13.613	21.587
3.351	5.329	11.413	22.607	64.470	111.550	8.588	14.892
1.704	2.596	5.522	15.538	22.062	47.578	3.516	6.024
0.996	1.164	0.174	7.226	1.294	3.946	1.107	1.673

Simulação do Modelo AR(1)
Resultados para:

Phi1 0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	5	30	198	44
0.5	5	21	180	54
1	4	12	147	28
1.5	4	4	57	15
2	2	2	6	6
3	1	1	1	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
3.911	6.709	23.932	36.608	168.283	227.277	30.889	57.051
3.828	6.512	15.776	26.324	150.398	210.382	38.788	69.132
2.976	5.384	7.706	16.794	116.302	177.038	18.574	37.266
2.346	4.714	1.862	6.798	34.016	80.704	9.998	20.082
1.217	1.863	0.592	2.788	-2.242	13.822	3.855	7.865
1.000	1.000	1.000	1.000	1.000	1.000	1.315	2.145

Simulação do Modelo AR(2)

Resultados para:

Phi1 0.1
Phi2 0.8

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	81	227	198	111
0.5	86	256	204	111
1	66	215	193	97
1.5	81	231	209	108
2	81	228	222	120
3	79	241	185	102

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
61.257	99.843	195.132	258.868	166.658	230.182	87.329	134.771
62.403	109.077	225.135	287.205	170.893	237.667	85.392	137.468
48.350	84.210	182.760	247.100	161.356	225.104	74.186	119.694
59.802	101.578	199.018	262.522	178.242	238.938	84.340	131.420
59.196	103.644	195.810	260.210	190.319	252.801	93.331	146.949
61.686	95.974	209.467	271.813	155.613	213.927	81.228	122.692

Simulação do Modelo AR(2)

Resultados para:

Phi1 0.3

Phi2 0.6

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	21	125	246	47
0.5	23	91	287	81
1	24	133	294	73
1.5	23	131	246	71
2	17	124	261	51
3	21	118	275	56

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
14.665	27.515	100.671	148.609	215.208	277.572	34.862	60.058
17.613	28.507	71.814	109.806	257.840	316.620	60.606	101.754
18.599	30.061	107.588	159.152	265.814	322.306	56.450	89.230
17.716	28.024	104.822	157.518	214.400	278.020	53.401	88.119
12.559	21.361	97.642	150.038	231.701	290.359	36.356	65.964
15.949	25.951	93.597	141.663	244.595	305.305	43.143	69.637

Simulação do Modelo AR(2)

Resultados para:

Phi1 0.5

Phi2 0.4

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	12	56	283	45
0.5	10	52	227	61
1	12	58	274	61
1.5	10	62	280	58
2	13	47	255	52
3	9	48	260	40

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
9.053	14.467	42.172	69.468	253.497	313.363	32.022	58.558
7.781	12.839	38.787	64.613	195.772	257.528	43.356	77.784
9.039	14.681	42.526	73.894	244.955	302.265	46.675	74.865
8.055	12.065	46.535	77.025	250.493	310.287	41.434	73.906
9.471	17.349	35.827	58.733	222.797	286.323	36.764	66.916
6.558	11.022	34.865	60.475	230.007	290.893	27.451	52.549

Simulação do Modelo AR(2)

Resultados para:

Phi1 0.7
Phi2 0.2

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	8	46	256	56
0.5	8	35	245	46
1	8	36	219	40
1.5	8	36	234	45
2	6	28	237	43
3	6	17	212	41

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
5.888	9.592	34.029	57.771	226.772	285.848	41.179	71.061
6.489	9.631	25.980	43.620	213.900	276.080	32.763	59.137
6.167	10.233	26.540	45.340	189.433	248.467	29.967	50.833
5.791	9.489	26.765	46.095	202.981	264.339	33.147	56.733
4.689	7.871	18.278	38.202	206.165	267.675	30.581	55.679
4.253	7.187	10.739	22.501	178.385	245.935	29.089	52.591

Simulação do Modelo MA(1)
Resultados para:

Theta1 -0.3

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	108	104	217	147
0.5	57	83	141	70
1	17	80	74	24
1.5	8	47	30	9
2	4	23	11	4
3	2	4	3	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
85.868	129.832	83.592	125.208	188.017	246.243	121.249	171.751
43.894	70.886	65.230	100.490	114.307	166.913	53.951	85.249
12.725	21.035	59.709	100.871	54.987	92.913	17.473	30.127
6.431	9.709	31.283	62.357	22.539	36.941	7.151	10.529
2.920	4.280	9.950	36.910	8.105	13.975	3.172	4.688
1.504	2.056	0.822	6.598	2.329	3.651	1.558	2.122

Simulação do Modelo MA(1)
Resultados para:

Theta1 -0.5

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	68	62	195	112
0.5	34	49	153	55
1	14	51	88	21
1.5	6	35	38	8
2	3	15	13	4
3	2	4	2	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
53.398	82.462	49.445	75.235	167.834	222.506	89.774	134.686
25.174	42.706	37.191	61.469	126.398	179.842	41.966	67.554
10.978	17.262	38.492	64.208	67.722	107.778	16.184	25.276
4.658	7.002	22.405	47.135	28.208	48.332	6.653	10.067
2.718	3.822	8.847	21.133	9.691	16.389	3.298	4.742
1.373	1.867	1.022	7.658	1.822	3.158	1.397	2.083

Simulação do Modelo MA(1)
Resultados para:

Theta1 -0.7

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	39	37	147	67
0.5	19	29	106	47
1	10	27	60	17
1.5	4	15	27	6
2	3	13	12	4
3	1	3	4	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
28.445	48.575	25.664	47.376	119.154	174.326	49.557	83.543
14.879	23.961	20.732	37.368	83.920	127.140	34.353	59.707
8.420	12.400	20.219	34.321	47.286	73.494	13.178	21.042
3.519	5.141	9.900	19.560	18.819	35.361	5.151	7.649
2.235	3.465	7.199	19.281	8.595	15.365	3.184	4.976
1.128	1.392	0.293	4.987	1.275	6.165	1.297	1.803

Simulação do Modelo MA(1)
Resultados para:

Theta1 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	32	38	166	95
0.5	25	36	129	41
1	9	27	54	17
1.5	5	24	23	7
2	2	22	16	4
3	1	4	2	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
25.621	38.419	30.663	45.497	139.418	193.522	74.086	116.234
19.400	30.240	26.995	44.845	104.196	152.864	31.932	50.928
7.469	10.551	19.760	33.420	40.684	68.076	13.690	20.810
3.874	5.686	13.157	34.443	17.535	28.465	5.377	8.123
2.065	2.895	13.422	31.018	11.589	20.311	3.027	4.873
1.201	1.499	1.634	5.946	1.696	3.124	1.412	1.868

Simulação do Modelo MA(2)

Resultados para:

Theta1 -0.3
Theta2 -0.1

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	80	87	205	103
0.5	61	98	169	82
1	23	68	83	29
1.5	9	60	52	11
2	4	35	13	4
3	2	4	4	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
61.150	98.570	68.127	106.653	174.487	234.553	81.566	124.714
48.271	74.409	75.653	119.567	141.801	196.299	65.903	99.017
17.874	28.566	49.944	85.516	64.586	101.114	22.895	35.945
7.197	11.203	41.451	77.949	37.669	66.551	8.528	12.872
3.107	4.553	20.562	49.198	9.406	16.414	3.641	5.319
1.441	1.999	1.552	5.728	1.942	5.518	1.505	2.075

Simulação do Modelo MA(2)
Resultados para:

Theta1 -0.3
 Theta2 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	97	113	264	135
0.5	50	100	192	63
1	25	112	119	29
1.5	10	83	54	12
2	4	67	26	5
3	2	19	4	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
75.614	119.206	87.703	137.897	235.034	293.466	109.593	159.447
37.374	63.386	76.270	122.750	161.428	222.452	47.842	78.798
18.640	30.720	88.109	136.451	94.075	144.865	21.735	35.365
7.414	13.126	59.599	106.581	35.662	71.958	8.728	14.592
3.213	4.847	44.164	90.756	17.333	35.607	4.008	5.952
1.391	1.889	5.899	32.421	2.660	4.800	1.414	1.926

Simulação do Modelo MA(2)

Resultados para:

Theta1 -0.7

Theta2 -0.1

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	37	42	154	75
0.5	19	32	138	48
1	9	24	66	16
1.5	4	22	29	7
2	3	16	16	4
3	1	7	3	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
28.722	44.398	30.512	54.188	127.096	181.464	59.484	90.996
13.682	24.158	22.746	40.434	111.106	165.454	33.750	61.950
6.866	10.414	16.847	30.193	49.703	82.257	12.966	19.894
3.203	4.457	14.826	29.534	20.937	37.983	6.022	8.458
2.358	3.682	10.151	21.489	10.931	21.289	3.139	4.981
1.237	1.663	2.160	12.760	2.010	4.150	1.486	2.034

Simulação do Modelo MA(2)
Resultados para:

Theta1 -0.7
 Theta2 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	26	37	223	74
0.5	20	34	175	48
1	9	37	135	19
1.5	4	24	69	9
2	3	22	34	4
3	1	3	6	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
18.679	33.521	27.322	47.298	191.367	253.953	55.842	92.638
14.479	25.301	26.030	42.630	145.605	203.895	35.659	59.621
7.033	11.447	26.091	48.849	106.126	164.134	15.081	22.959
3.661	5.299	18.113	29.927	50.629	87.891	6.853	11.027
2.228	3.532	13.702	30.658	22.349	44.831	3.386	5.054
1.143	1.477	1.088	5.072	2.492	8.948	1.429	1.911

Simulação do Modelo MA(2)
Resultados para:

Theta1 -0.9
 Theta2 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	18	37	181	68
0.5	14	30	156	37
1	7	32	95	19
1.5	4	18	43	7
2	2	15	23	4
3	1	2	2	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
13.722	21.478	28.656	45.024	150.556	210.724	50.398	84.962
11.528	17.332	22.183	36.957	128.407	183.233	26.743	47.737
5.695	8.645	23.215	40.045	73.353	117.567	14.759	23.541
3.136	4.844	11.887	23.993	30.370	54.770	5.886	8.594
1.736	2.544	10.181	18.919	15.050	30.310	3.139	4.801
1.037	1.303	1.034	3.506	0.951	3.109	1.281	1.839

Simulação do Modelo ARMA(1,1)

Resultados para:

Phi1 0.1
Theta1 -0.1

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	137	124	196	149
0.5	70	122	160	93
1	27	106	59	27
1.5	10	73	22	10
2	5	44	10	5
3	2	15	3	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
110.249	163.971	97.266	150.294	166.060	225.820	120.757	176.723
53.233	87.167	97.491	147.229	131.618	187.502	72.463	114.197
17.985	35.715	81.666	131.174	42.435	75.445	18.299	35.981
6.863	12.477	49.248	96.352	17.030	26.910	7.172	12.808
4.112	5.968	26.373	62.267	7.254	11.966	4.432	6.408
1.594	2.206	2.651	26.849	2.238	3.582	1.700	2.360

Simulação do Modelo ARMA(1,1)
Resultados para:

Phi1 0.1
 Theta1 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	28	31	128	68
0.5	17	30	106	41
1	9	21	57	12
1.5	4	27	35	8
2	2	17	15	4
3	1	6	3	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
22.280	32.840	25.312	37.208	103.839	152.581	52.019	84.101
14.001	20.399	23.065	37.715	83.612	127.648	32.048	49.252
6.949	10.651	15.068	27.392	42.033	72.567	9.627	14.653
2.860	4.460	19.528	33.912	25.865	43.475	5.850	9.470
1.979	2.861	9.757	24.103	10.060	19.760	3.194	4.906
1.129	1.531	2.600	8.480	1.634	4.166	1.379	1.901

Simulação do Modelo ARMA(1,1)

Resultados para:

Phi1 0.5
Theta1 -0.1

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	28	36	187	66
0.5	20	35	181	40
1	10	29	93	19
1.5	5	31	60	10
2	3	21	26	4
3	1	4	5	1

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
21.954	33.766	26.122	45.038	157.296	216.604	48.808	82.972
14.800	25.760	24.764	45.116	150.244	211.216	30.337	49.543
8.034	12.646	21.864	36.796	71.005	114.055	14.893	22.967
3.925	6.375	21.205	40.355	46.422	73.438	7.298	11.822
2.267	3.373	13.656	28.344	17.093	34.907	3.039	4.601
1.070	1.310	0.651	8.049	2.313	7.347	1.225	1.615

Simulação do Modelo ARMA(1,1)

Resultados para:

Phi1 0.9
 Theta1 -0.9

ARL

K	Shewhart	EWMA	AR(1)	Wheeler
0	5	34	39	47
0.5	3	26	27	50
1	3	7	9	31
1.5	2	1	1	12
2	1	1	1	10
3	1	1	1	2

Intervalo de Confiança (95%)

Shewhart		EWMA		AR(1)		Wheeler	
min	max	min	max	min	max	min	max
4.066	5.914	25.935	41.765	29.986	47.614	35.605	58.035
2.061	3.219	18.851	33.649	19.033	35.947	33.196	65.944
2.121	3.639	2.838	11.122	4.153	13.967	20.464	41.276
1.192	2.408	1.000	1.000	1.000	1.000	6.765	16.355
1.008	1.312	1.000	1.000	1.000	1.000	3.584	16.456
0.950	1.410	1.000	1.000	1.000	1.000	1.317	2.603