

O ambiente R como uma proposta de renovação para aprendizagem e monitoramento de processos em Controle Estatístico de Qualidade

Elisa Henning (UFSC) dma2eh@joinville.udesc.br

Custodio da Cunha Alves (UNIVILLE) custodio.alves@univille.br

Vilson Vieira (UDESC) vilson@void.cc

Resumo: Este artigo tem como objetivo apresentar o ambiente R como uma proposta de ferramenta computacional para o ensino de outras técnicas de monitoramento e Controle Estatístico de Qualidade. Para apresentação deste ambiente utilizam-se duas alternativas muito eficazes e complementares ao gráfico de controle Shewhart para sinalizar com maior rapidez pequenos desajustes: os gráficos de controle com memória CUSUM e EWMA. Comumente são utilizados pacotes comerciais para a resolução de exercícios nesta área, muitas vezes exemplificados e recomendados pelos próprios autores em seus livros. O diferencial está no fato do R ser um projeto GNU baseado no conceito de software livre. Apresenta-se aqui um breve panorama das aplicações do R para Controle Estatístico de Qualidade para os gráficos citados, através da utilização do pacote qcc. O exemplo utilizado é simples sendo de fácil entendimento pelos acadêmicos de cursos de graduação e pós-graduação. Conclui-se que o R é uma ótima ferramenta alternativa para esta área. Necessita porém do desenvolvimento de material didático e auxiliar específico para este fim. Palavras-chave: Ambiente R; Ensino; Controle Estatístico de Qualidade.

1. Introdução

A utilização de pacotes computacionais tem sido cada vez mais freqüente no meio acadêmico. Não poderia ser diferente no campo do ensino de Controle Estatístico de Qualidade que acompanha os avanços da tecnologia para atender a complexidade dos atuais processos. Houve grande evolução nesta área, na medida em que os pacotes permitiram aos acadêmicos fixarem-se mais na análise e interpretação de resultados. Diversos livros recomendam e trazem exemplos e exercícios resolvidos com auxílio destas ferramentas.

Atualmente vários pacotes são encontrados para Controle Estatístico de Processos tais como os pacotes MINITAB® e SAS® que têm recursos de controle de qualidade atrativos (DEVORE, 2000). Como característica comum está o fato de todos serem pacotes comerciais, implicando no pagamento de licenças. Não raras vezes este custo é alto para os acadêmicos dos cursos tanto de nível de graduação quanto de pós-graduação. Felizmente, hoje, muitos pacotes livres já podem ser encontrados e, dentre eles, o ambiente R.

O R (Ihaka & Gentleman, 1996) é uma linguagem e um ambiente para computação estatística que traz muitas vantagens ao usuário. É um projeto GNU baseado no conceito de software livre podendo ser usado sem custos de licença (VENABLES, 2003). Ele permite, de acordo com os autores, uma grande variedade de análises estatísticas como, por exemplo, análise exploratória de dados, testes estatísticos, regressão linear e não linear, análise de séries temporais, entre muitas outras. Um dos pontos fortes deste pacote é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário. Além disso, existem inúmeras funções para manipulação, importação e exportação de dados. É multi-plataforma, contendo versões para Windows, MacOS, GNU/Linux e Unix (BEASLEY, 2004). O termo ambiente pretende caracterizar o R como um sistema totalmente planejado e

coerente ao invés de uma aglomeração de ferramentas muito específicas e inflexíveis, permitindo aos usuários acrescentar funcionalidade adicional por definição de novas funções (BEASLEY, 2004). O autor ainda cita uma lista de recursos de ajuda na *home-page* do projeto, destacando-se o *r-help*, uma lista mais apropriada para os usuários.

O R pode ser baixado diretamente da Internet em página específica: www.r-project.org. Além disso, existem ainda mais de 700 pacotes que estendem R, acrescentando-lhe diversas funcionalidades, indo desde a implementação de métodos específicos de análise estatística, até pacotes para a criação de interfaces gráficas com o usuário. Estes pacotes estão disponíveis livremente em um repositório central, o CRAN (Comprehensive R Archive Network). Além de usufruir tais pacotes é possível criar novos utilizando a própria linguagem R ou outra linguagem de programação externa como C por exemplo.

Este documento irá apresentar brevemente pacote QCC (Quality Control Charts) e algumas de suas principais aplicações voltadas aos Gráficos de Controle de Soma Acumulada (CUSUM) e da Média Móvel Exponencialmente Ponderada (EWMA).

2. Aplicações do R ao Controle Estatístico de Qualidade

2.1 O pacote qcc

O pacote qcc (Quality Control Charts) foi desenvolvido por Lucca Scrucca do Departamento de Ciências Estatísticas da *Università degli Studi di Perugia*, Itália. O autor teve a idéia de desenvolver um pacote computacional para proporcionar aos seus alunos de graduação uma ferramenta para a aprendizagem dos conceitos básicos de Controle Estatístico de Qualidade (SCRUCCA, 2004). Por ser um ambiente GNU, o R foi um candidato natural. No seu desenho inicial as funções eram similares às do pacote comercial S-PLUS. No entanto, o processo de desenvolvimento do pacote permitiu que as versões mais novas tivessem características próprias, embora algumas semelhanças como o S-PLUS ainda sejam encontradas (SCRUCCA, 2004). O pacote qcc do R permite: plotar gráficos de controle Shewhart para variáveis e atributos, desenhar Curvas Características de Operação, efetuar a Análise da Capacidade Processos, construir gráficos de Pareto e diagramas de Causa e Efeito e plotar Gráficos de Controle da Soma Acumulada (CUSUM) e Média Móvel Exponencialmente Ponderada (EWMA) (SCRUCCA, 2004). O pacote pode ser baixado diretamente da página do projeto, tanto para Linux como Windows.

Na figura 1 mostra-se o acesso ao pacote qcc na tela de trabalho do R e este disponibiliza uma ajuda on-line que pode ser visualizada na figura 2. É possível também obter uma versão para impressão em formato pdf. As referências utilizadas para as rotinas estão na ajuda citada.

2.2 Gráfico de Controle da Soma Acumulada (CUSUM)

Os gráficos de Controle de Shewhart apresentam uma desvantagem que é a falta de habilidade para detectar uma mudança relativamente pequena na média do processo (DEVORE, 2000 pág 624). Isto se deve ao fato de usar apenas a informação do processo contida no último ponto plotado, ignorando seqüência inteira de pontos (MONTGOMERY, 2004). Os gráficos de controle e procedimentos de soma acumulada (CUSUM) foram projetados, segundo Devore (2000), para corrigir esta lacuna. Há duas versões equivalentes de um procedimento CUSUM para a média do processo, uma gráfica e a outra computacional, o Cusum tabular ou algorítmico. Ainda de acordo com Montgomery (2004), é possível planejar procedimentos de soma acumulada para outras variáveis com objetivo da modelagem de não-conformidades e falhas do processo. Este documento refere-se apenas ao gráfico da soma acumulada para a média do processo.

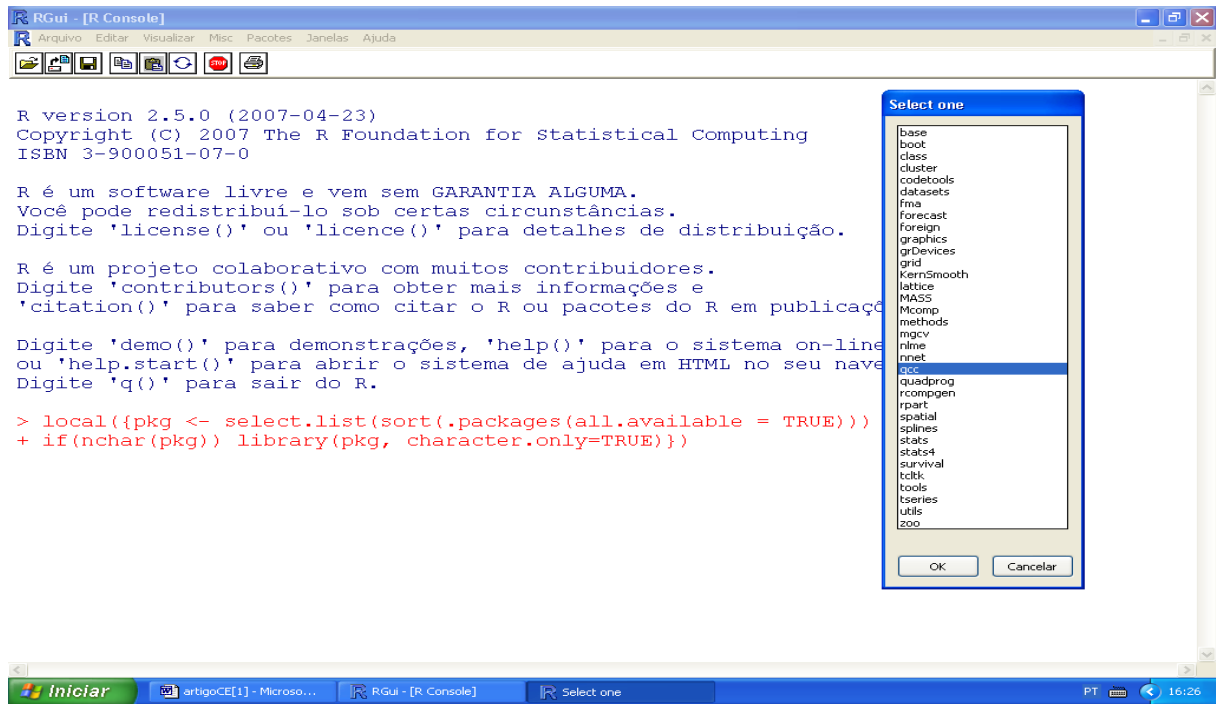


Figura 1 - Tela inicial do R (no Windows).
Fonte: Elaborada pelos autores.

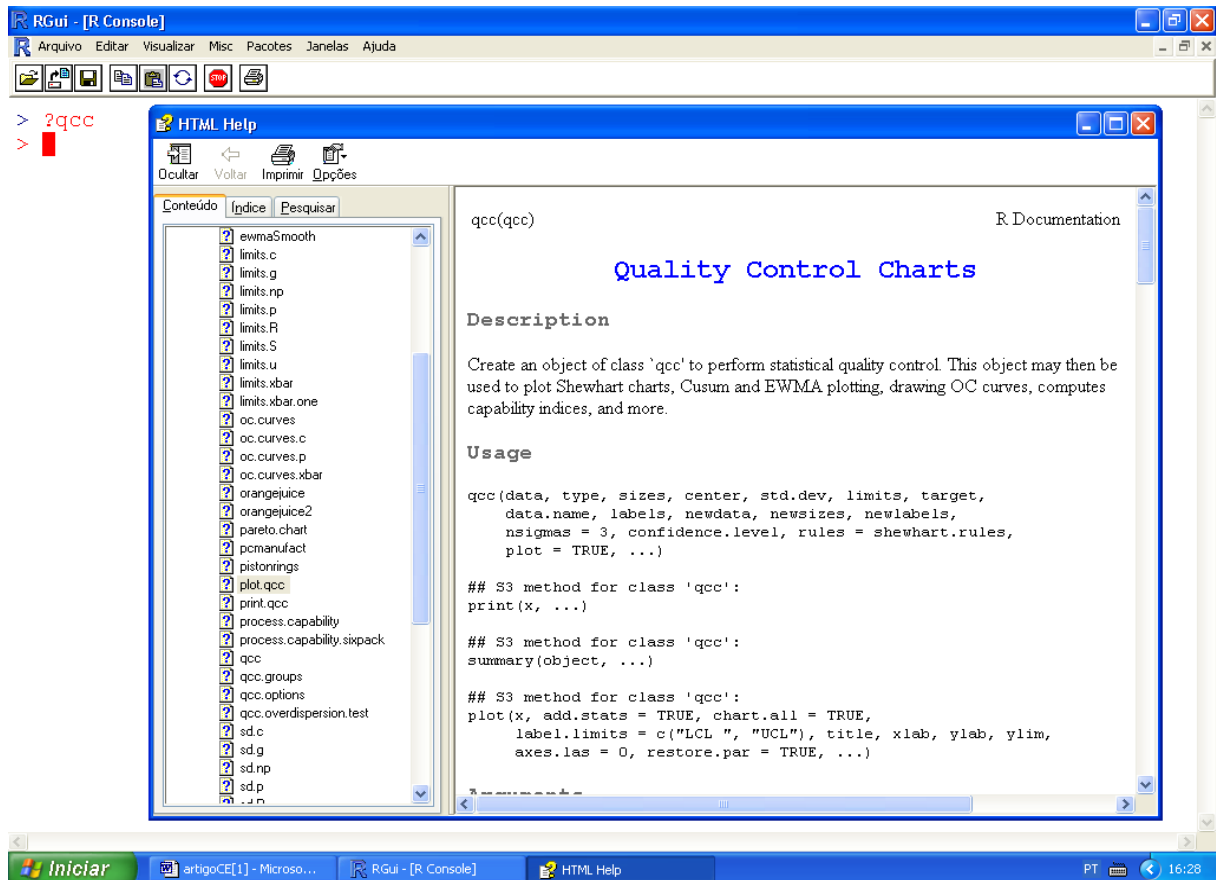


Figura 2 - Ajuda do R.
Fonte: Elaborada pelos autores.

Segundo Montgomery (2004), os procedimentos de soma acumulada podem ser utilizados tanto para observações individuais quanto para subgrupos racionais. Estes procedimentos trabalham acumulando desvios da média que estão acima ou abaixo do valor alvo μ_0 (nominal), com as estatísticas C^+ e C^- denominadas Cusum unilateral superior e inferior, respectivamente. Ainda, segundo o autor são calculados da seguinte forma:

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

$$C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-] \quad (2)$$

Nas equações acima, (1) e (2), o valor de K é comumente chamado de valor de referência e é escolhido a meio caminho entre o valor alvo μ_0 e o valor da média fora de controle μ_1 que se deseja detectar com rapidez (MONTGOMERY, 2004). Deste modo se a mudança é expressa em termos de desvio padrão então K pode ser considerado como a metade da magnitude da mudança. As estatísticas C^+ e C^- acumulam desvios a partir do valor alvo que são maiores do que K, com abas as quantidades recolocadas em zero ao tornarem-se negativas. Caso C^+ ou C^- excederem um intervalo de decisão H, o processo é considerado como fora de controle. Um valor razoável para H costuma ser cinco vezes o desvio padrão do processo, σ (MONTGOMERY, 2004). A construção do gráfico é feita plotando-se os valores de C^+ e C^- versus o número da amostra, além das observações x_i para cada período. Assim, é possível ao usuário do gráfico de controle visualizar o desempenho do processo que o levou a um valor específico de cusum (MONTGOMERY, 2004).

2.3 Gráfico de Controle da Média Móvel Exponencialmente Ponderada (EWMA)

Para Montgomery (2004), o gráfico de controle da média móvel exponencialmente ponderada (EWMA) é também uma boa alternativa ao gráfico de controle Shewhart, quando estamos interessados na detecção de pequenas mudanças. Seu desempenho é aproximadamente equivalente ao do gráfico de controle das somas acumuladas, e é de certa forma, de construção e operação mais simples. Do mesmo modo como no caso do CUSUM, é usualmente empregado para observações individuais, mas também pode ser usado para subgrupos racionais. O gráfico de controle de EWMA é definido, de acordo com o autor, como:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (3)$$

onde $0 < \lambda \leq 1$ é uma constante e seu valor inicial z_0 é o alvo μ_0 do processo, podendo também ser usada a média de dados preliminares em alguns casos. Como o EWMA pode ser considerado como uma média ponderada de todas as observações anteriores com as atuais, este gráfico torna-se insensível à hipótese de normalidade, sendo considerado por Montgomery (2004) como o ideal para observações individuais.

O gráfico é construído plotando-se os valores z_i versus o número i da amostra. A linha central corresponde ao valor alvo μ_0 . Segundo Montgomery (2004), os limites superior e inferior de controle para este gráficos são os seguintes:

$$LSC = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]} \quad (4)$$

$$LIC = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]} \quad (5)$$

O fator L corresponde à largura dos limites de controle. Usualmente valores de λ entre 0,05 e 0,25 funcionam bem na prática de acordo com o autor. Este também considera uma boa

regra empírica usar menores valores de λ para detectar menores mudanças. Já para a escolha de L , este, com o valor de três ($L=3$), ou seja os limites três sigmas usuais, funciona bem com valores maiores de λ . Para valores de λ menores o autor sugere reduzir a largura dos limites utilizando um valor de L menor.

2.4 Gráficos de Soma Acumulada com o R

O R permite somente construir cusum para variáveis contínuas. Inicialmente é necessário que se tenha um objeto “qcc”, ou seja, deve-se construir um gráfico X_i por exemplo. A sintaxe completa pode ser vista no quadro 01.

```
cusum(object, decision.int = 5, se.shift = 1,
      label.bounds = c("LDB", "UDB"), add.stats = TRUE, chart.all =
      TRUE, ylim = NULL, axes.las = 0, restore.par = TRUE, ...)
```

Quadro 01 Sintaxe do R para o gráfico cusum.

Fonte: Elaborado pelos autores.

O R constrói o gráfico cusum com os valores de C^+ e C^- , o número da amostra ou grupo e os valores das observações individuais ou médias dos grupos (SCRUCCA, 2004). O valor alvo considerado é a média das observações ou grupos. É possível escolher um outro valor para tal. A função *target*, quando da construção do gráfico de Shewhart permite esta opção. De acordo com Scrucca (2004) é considerado o valor de $H = 5$ para o intervalo de decisão, sendo também permitido outro valor, alterando-se o valor em *decision.int*. O tamanho da mudança a ser detectado é igual a 01 (um) σ , permitindo também outros valores modificando-se este na função *se.shift*. Além do gráfico, o R retorna o número de grupos ou observações, o valor alvo, o desvio padrão, o intervalo de decisão, o tamanho da mudança e o número de pontos além dos limites de controle (SCRUCCA, 2004). Na figura 04 tem-se um exemplo da saída do R. Os dados utilizados 30 observações individuais, sendo que as 20 primeiras foram extraídas de uma distribuição normal com média $\mu = 10$ e desvio padrão $\sigma = 1$ e, para as dez últimas observações considerou-se uma distribuição normal todavia com média $\mu = 11$ e desvio padrão $\sigma = 1$. Pode-se considerar neste caso que estas dez últimas observações foram extraídas do processo quando ele estava fora de controle, ou seja depois do processo ter sofrido uma mudança de 1 (um) σ na média. Este é um exemplo extraído de Montgomery (2004, pág. 256) cujos dados estão dispostos na tabela 1.

Tabela 1: Dados para o Exemplo

Observação i	x_i	Observação i	x_i	Observação i	x_i
1	9,45	11	9,03	21	10,9
2	7,99	12	11,47	22	9,33
3	9,29	13	10,51	23	12,29
4	11,66	14	9,40	24	11,50
5	12,16	15	10,08	25	10,60
6	10,18	16	9,37	26	11,08
7	8,04	17	10,62	27	10,38
8	11,46	18	10,31	28	11,62
9	9,20	19	8,52	29	11,31
10	10,34	20	10,84	30	10,52

Fonte: Montgomery (2004, pág. 256).

As observações foram plotadas em gráficos de controle de Shewhart e CUSUM para observações individuais cuja saída do R é mostrado nas figuras 03 e 04.

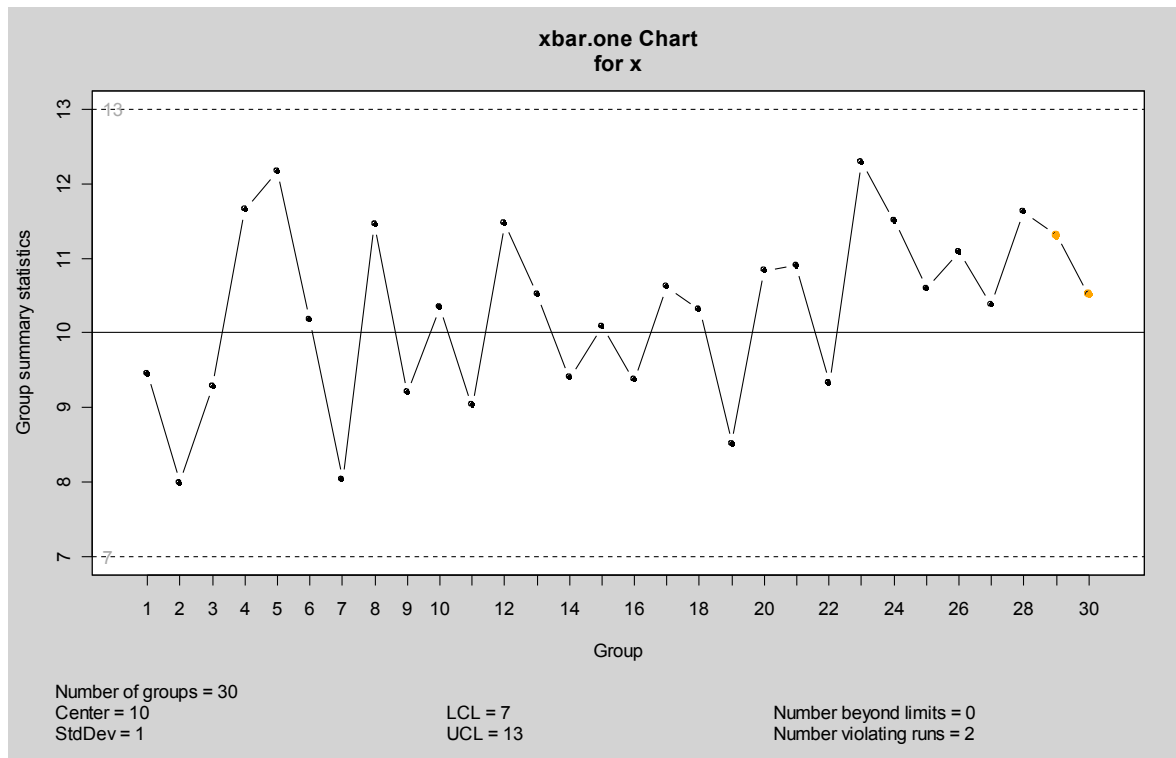


Figura 03 - Gráfico de Shewhart (Observações Individuais) plotado pelo R para os dados da Tabela 1.
Fonte: Elaborado pelos autores.

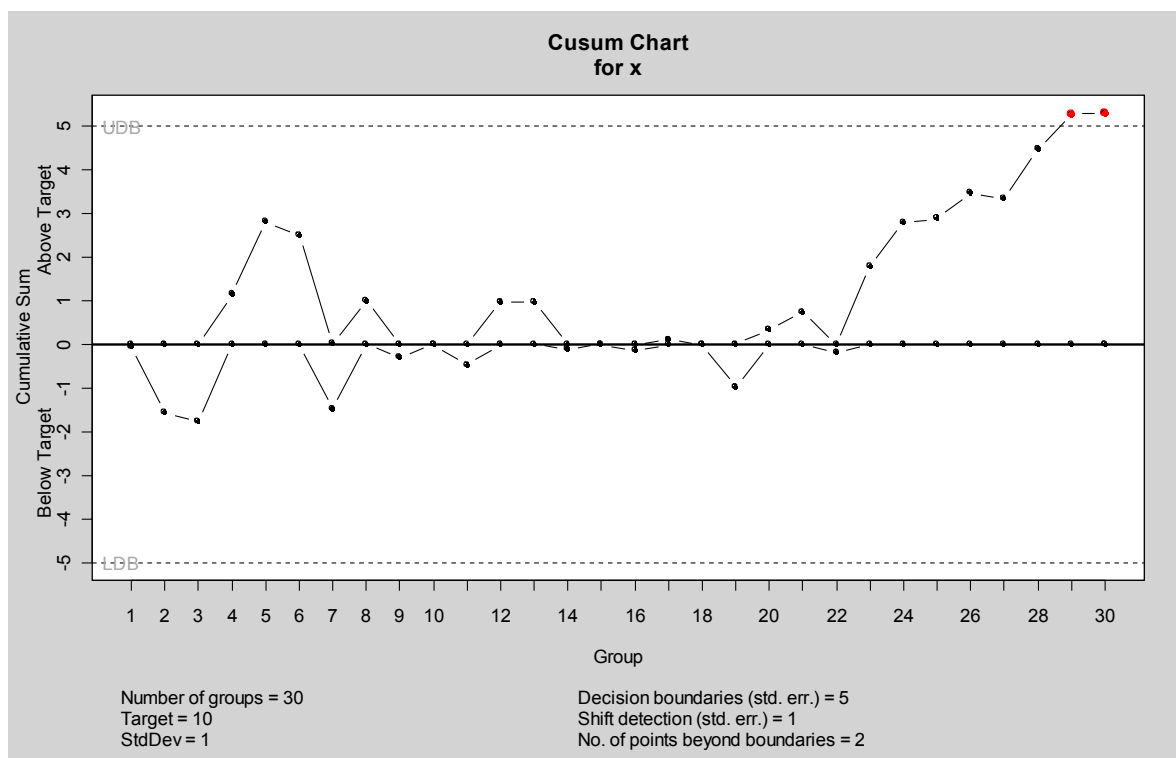


Figura 04 - Gráfico CUSUM plotado pelo R para os dados da Tabela 1.
Fonte: Elaborado pelos autores.

O gráfico e os valores retornados para o CUSUM estão de acordo com o exposto em

Montgomery (2004). Como nenhum dos pontos no gráfico de Shewhart (figura 03) se localiza além dos limites de controle, não se tem evidências de que o processo esteja fora de controle. Percebe-se que em se utilizando o sinal tradicional de um processo fora de controle, o gráfico falha em detectar a mudança na média. A razão para esta falha está, de acordo com o autor na magnitude relativamente pequena da mudança. Como o R utiliza as demais regras sensibilizantes tem-se uma indicação de uma mudança no nível do processo para os últimos pontos. O gráfico cusum (figura 04) indica que esta mudança provavelmente entre os períodos 22 e 23, pois no período 22 foi a última vez que o processo esteve sob controle. A daí tem-se apenas valores positivos para o CUSUM.

Para plotar o gráfico de Shewhart para observações individuais foram estabelecidos os valores centrais (10), limite de controle superior (13), limite de controle inferior (7) e desvio padrão igual a 1 (um). Para o cusum, considerou-se $H = 5$ e o tamanho da mudança igual a 1σ ($K = \frac{1}{2}$).

2.5 Gráfico de Controle da Média Móvel Exponencialmente Ponderada (EWMA) com R

Analogamente ao gráfico CUSUM, a construção do EWMA necessita de um objeto qcc. A função `ewma(qcc)` constrói o EWMA. A sintaxe completa pode ser vista no quadro 2.

```
ewma(object, lambda=0.2, nsigmas=object$nsigmas,
      add.stats = TRUE, xlab, ylab, ylim = NULL, axes.las = 0,
      restore.par = TRUE, ...)
```

Quadro 02 - Sintaxe do R para EWMA.

Fonte: Elaborado pelos autores.

A figura 05 mostra a saída do R para os dados para o exemplo citado.

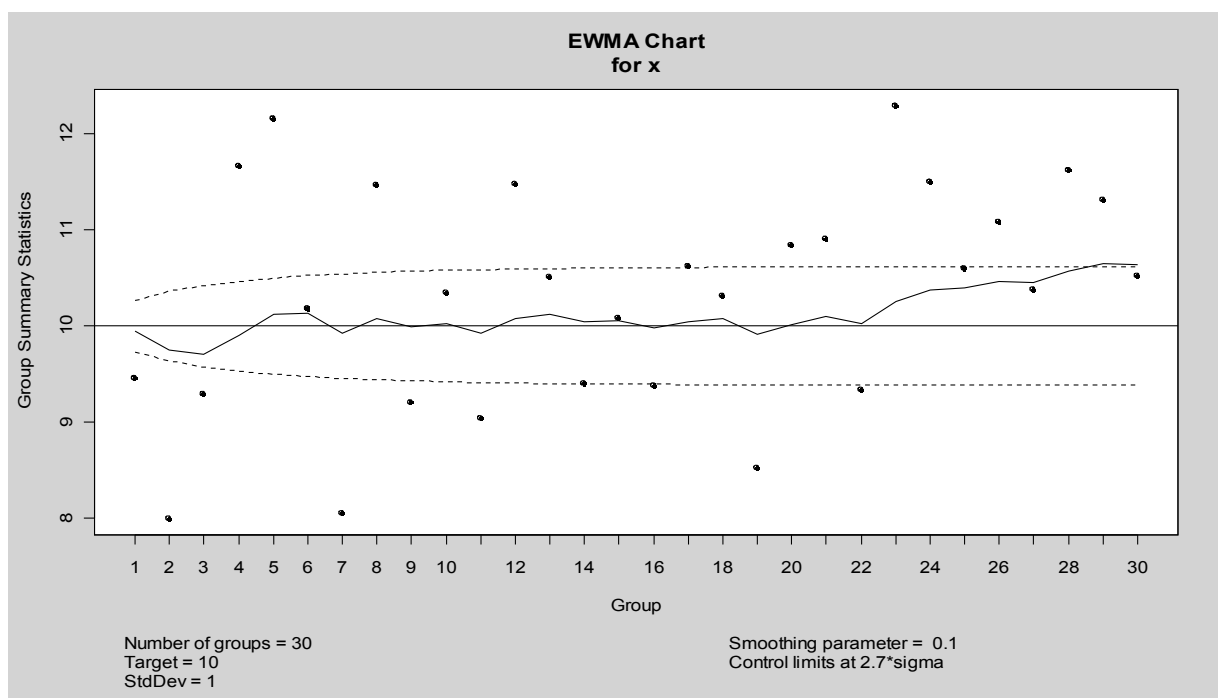


Figura 05 - Gráfico EWMA para os dados da Tabela 1 com o R.

Fonte: Elaborado pelos autores.

O R retorna o número de grupos ou observações, o valor alvo, o desvio padrão do

processo, o valor do parâmetro de suavização λ e os limites de controle. O R considera como valor alvo a média do processo e limites de controles iguais a 3 sigmas, permitindo mudanças nestes. O MMEP depende de um valor inicial de λ , que no default é considerado igual a 0,2, podendo ser modificado através do argumento *lambda*.

O gráfico (figura 05) e valores estão de acordo com Montgomery (2004), sinalizando nas últimas observações, portanto, pode-se concluir que o processo está fora de controle. Para este caso foi considerado $L = 2,7$ e $\lambda = 0,1$, conforme o exemplo de Montgomery (2004, pag 270).

2.6 Investigação de Normalidade dos Dados

Ao utilizarmos gráficos de controle para observações individuais é importante verificar a hipótese de normalidade (MONTGOMERY, 2004). Esta verificação pode ser realizada também em ambiente R que plota tanto o gráfico de probabilidade normal quanto alguns testes formais tais como o Shapiro-Wilk e o teste Kolmogorov-Smirnov. A verificação para os dados utilizados no exemplo é mostrado na figura 06 (gráfico de probabilidade normal) e quadro 03 (teste Shapiro-Wilk).

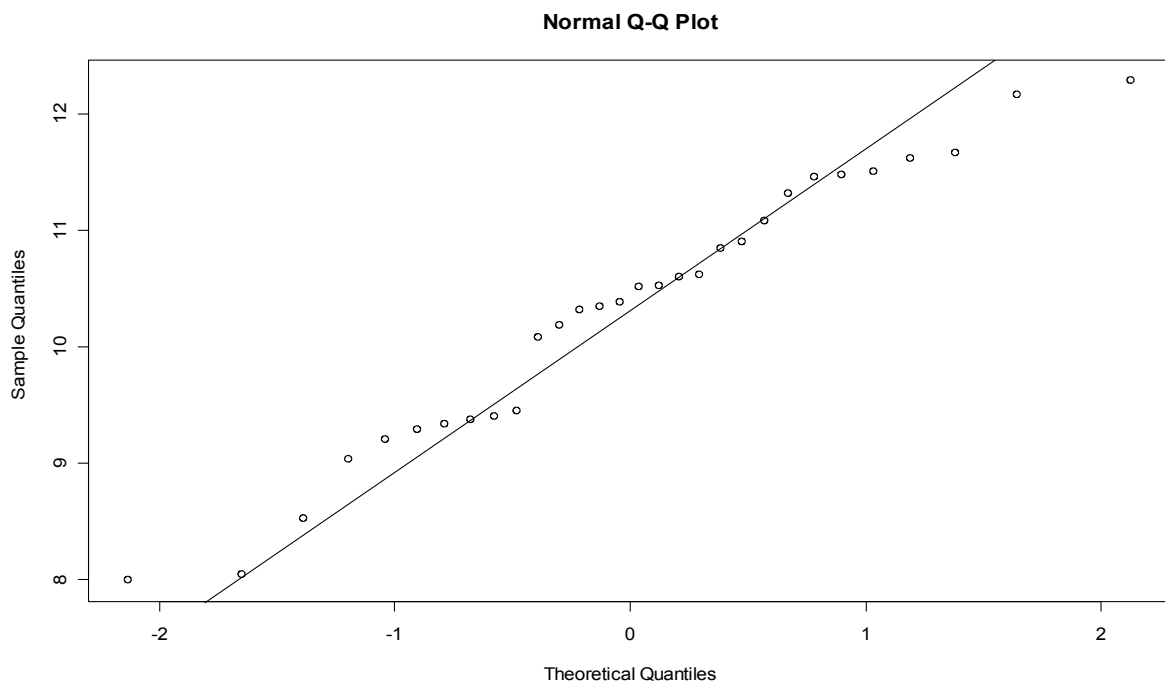


Figura 06 - Gráfico de probabilidade normal com o R para os dados da Tabela 1

Fonte: Elaborado pelos autores.

```
> shapiro.test(x)

      Shapiro-Wilk normality test

data:  x
W = 0.9664, p-value = 0.4454
```

Quadro 03 - Teste Shapiro Wilk para os dados da Tabela 1.

Fonte: Elaborado pelos autores.

Outros testes para verificação de normalidade também podem ser executados pelo R,

através de pacotes adicionais. O pacote fBasics (Wuertz, 2006), por exemplo, inclui os testes Jarque Bera, Lilliefors e Anderson Darling, entre outros.

3. Conclusões e Considerações Finais

Pelo exposto anteriormente pode-se concluir que o ambiente R é uma excelente alternativa prática e econômica para auxiliar o acadêmico na construção de gráficos CUSUM e EWMA. A execução destas rotinas específicas para a construção destes gráficos permite otimizar o tempo para a escolha adequada do gráfico de controle a ser utilizado.

A principal vantagem está no fato de ser um ambiente computacional livre, podendo ser baixado sem custos de licença diretamente da Internet. Mas as aplicações futuras são ilimitadas. Pelo fato de ser de código aberto é possível vislumbrar o desenvolvimento de gráficos CUSUM e EWMA para modelagem de não-conformidades e falhas do processo, por exemplo.

Uma lacuna é a pouca quantidade de material de apoio aos acadêmicos, principalmente em língua portuguesa especificamente para o ensino na engenharia apesar de alguns livros didáticos já apresentarem exemplos resolvidos com pacotes comerciais conforme abordado anteriormente. Material didático referente aos gráficos Shewhart, pode ser encontrado em <http://www.void.cc/r> (HENNING e VIEIRA, 2006).

A interface do R pode não ser muito atrativa ao acadêmico não familiarizado com alguma linguagem de programação pelo fato de ele ter de digitar os comandos. Pode-se estudar a viabilização do pacote qcc no Rcmdr que é uma interface gráfica mais amigável do R (FOX, 2004).

O pacote qcc do R contém alguns conjuntos de dados extraídos da obra de Montgomery que tornam ainda mais simples o trabalho do professor. Outras vantagens podem ser listadas, de acordo com Beasley (2004), como a facilidade para manipulação e armazenagem de dados, a inclusão de um conjunto de operadores para cálculo, em particular as matrizes, uma grande e coerente coleção integrada de ferramentas intermediárias para a análise de dados e facilidades gráficas para a análise de dados e visualização na tela ou na impressão. O ambiente R apresenta funções específicas para leitura de dados. É possível através dele, importar dados de diversos formatos, como por exemplo, em txt, csv, do EXCEL, do Minitab, do SAS, de banco de dados, entre outros. No entanto, exige-se um certo cuidado na apresentação dos dados, mas nada que complique muito o usuário. Os gráficos de controle podem ser salvos também em diversos formatos como Metafile, Postscript, pdf, jpeg, e bitmap. Existe também a possibilidade de rodar o R dentro do editor Emacs o que pode tornar mais simples a confecção de documentos. O gráfico de controle CUSUM tanto padronizado quanto para escala pode ser facilmente construído onde as transformações solicitadas podem ser obtidas através de simples operações entre vetores e escalares.

Referências

- BEASLEY, C. R. Bioestatística usando R. Universidade Federal do Pará. 2004. Disponível em: <http://www.r-project.org/other-docs.html>. Acessado em 01/09/2005.
- DEVORE, J. L. Probabilidade e Estatística para Engenharia e Ciências. São Paulo: THOMSON, 2006.
- FOX, John. The R Commander: A Basics-Statistics Graphical User Interface to R. Journal of Statistical Software. Volume 14. Disponível em: <http://www.jstatsoft.org/v14/i09/v14i09.pdf>. Acessado em 08/05/2007.
- IHAKA, R. GENTLEMAN, R. R. A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5 (3): pag 299-314.
- MONTGOMERY, D. C. Introdução ao Controle Estatístico de Qualidade. 4. ed. Rio de Janeiro: LTC, 2004.
- HENNING, E. VIEIRA, V. Controle Estatístico de Processos com o R. Disponível em: <http://www.void.cc/r>

Acessado em 01/08/2007.

SCRUCCA, L. Qcc: An R package for quality controle charting and statistical process control. R News, Vol 4/1: pag 11-17. Disponível em: [http:// www.r-project.org](http://www.r-project.org). Acessado em 11/08/2005.

SCRUCCA, L. The Qcc Package: Reference Manual. Disponível em: [http:// www.r-project.org](http://www.r-project.org). Acessado em 11/08/2005.

VENABLES, W. N, SMITH, D.M. An introduction to R. Disponível em: <http://www.r-project.org>. Acessado em 10/09/2004.

WUERTZ. D. The fBasics Package. Disponível em: <http://rmetrics.org>. Acessado em 08/05/2007.