



UMA BREVE VISÃO SOBRE APLICAÇÕES EM AMBIENTE R PARA O ENSINO DE ENGENHARIA

Elisa Henning – dma2eh@joinville.udesc.br

Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção
Campus Universitário Trindade

88040-900 – Florianópolis – Santa Catarina

Custodio da Cunha Alves – custodio@univille.net

Universidade da Região de Joinville, Departamento de Engenharia de Produção
Campus Universitário s/n

CEP 89201-974 – Joinville – Santa Catarina

Andréa Cristina Konrath – andreack@gmail.com

Universidade Federal do Rio Grande, Instituto de Matemática, Estatística e Física
Av. Itália km 08 - Pav. J - sala 25

96202-900 - Rio Grande - Rio Grande do Sul

Robert Wayne Samohyl – samohyl@yahoo.com

Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção
Campus Universitário Trindade

88040-900 – Florianópolis – Santa Catarina

***Resumo:** Este documento apresenta e faz uma breve discussão de alguns pacotes e funções do ambiente GNU R para o ensino de engenharia. O objetivo principal é a divulgação, principalmente no meio acadêmico. Aplicações direcionadas ao ensino de Controle Estatístico de Qualidade e Métodos de Previsão em Engenharia são sugeridas. Além disso, como complementação, é apresentada uma página na Internet para atender tal finalidade.*

***Palavras-chave:** Ambiente R, Controle Estatístico de Processo, Métodos de Previsão em Engenharia, Ensino.*

1 INTRODUÇÃO

Como a maioria das atividades práticas de engenharia envolve manipulação de uma grande quantidade de dados, o conhecimento e a prática de estatística computacional são imprescindíveis para a formação deste profissional. Métodos estatísticos, por exemplo, auxiliam o engenheiro no desenvolvimento de novos produtos, processos e sistemas, melhorando os existentes. Estes métodos são poderosas ferramentas que ajudam na compreensão da variabilidade presente nos problemas de engenharia (MONTGOMERY & RUNGER, 2003).

Portanto, o ensino desses métodos merece atenção especial. A utilização de pacotes computacionais como recurso de aprendizagem é fundamental para o ensino de estatística nos cursos de engenharia. Felizmente, uma grande evolução no desenvolvimento de pacotes

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



computacionais tem possibilitado aos acadêmicos aplicarem seus conhecimentos na análise estatística e interpretação de resultados. Além disso, muitos livros didáticos recomendam e trazem exemplos e exercícios resolvidos com auxílio destas ferramentas. Atualmente, vários pacotes comerciais com recursos atrativos estão disponíveis. Alguns pacotes livres podem ser encontrados e, dentre eles, o GNU R (R CORE DEVELOPMENT TEAM, 2008), objeto de estudo deste trabalho. O R é considerado por Silva (2006) um recurso inovador e fundamental para o ensino de estatística.

Fundamentado nesta evolução, torna-se o objetivo principal deste trabalho, divulgar o ambiente R para profissionais e alunos de engenharia. Pretende-se apresentar o R como uma ferramenta poderosa e acessível, para auxiliar na utilização de métodos estatísticos, tanto no contexto de sala de aula, quanto de pesquisa. Embora muito material esteja disponível, principalmente na Internet, a grande maioria foi escrito em inglês. Assim, procurou-se também reunir e escrever documentos em português, que vão desde instruções de instalação, comandos básicos, estatística básica e algumas aplicações específicas. Destacam-se dois temas de suma importância para a área de Engenharia que são: o Controle Estatístico de Processo e Métodos de Previsão em Engenharia. Para facilitar a utilização destes recursos desenvolveu-se uma página na Internet que disponibiliza tais documentos.

Este artigo está assim estruturado: uma breve introdução sobre a utilização de recursos computacionais no ensino de engenharia, um comentário sobre a aplicação do ambiente R é realizado na seção 2; a seção 3 apresenta brevemente os pacotes voltados para Controle Estatístico de Qualidade; os pacotes voltados para métodos de previsão em engenharia são descritos na seção 4 e, na seção 5 é apresentada uma página na Internet desenvolvida para divulgação de material direcionado ao ensino de engenharia. Na seção 6 estão algumas complementações que abordam pacotes úteis no ensino de estatística para acadêmicos dos cursos de engenharia. Finalizando, na seção 7, estão as conclusões e considerações finais.

2 PACOTE GNU R

O GNU R é uma linguagem e um ambiente para computação estatística. Como é um projeto GNU, baseado no conceito de software livre, pode ser usado sem custos de licença. Ele permite 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 diversas outras. Como ponto forte está a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos. Além disso, existem inúmeras funções para manipulação, importação e exportação de dados, sendo multi-plataforma, contendo versões para Windows, MacOS, GNU/Linux e Unix. É possível carregar dados externos das mais diversas formas, incluindo de planilhas eletrônicas, como o Microsoft Excel® e Open-Office, banco de dados e até de outros pacotes como MINITAB® e SPSS®, por exemplo. 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 R pode ser obtido diretamente da Internet em sítio específico: <http://www.r-project.org>. Para complementar, coloca-se que, existem mais de 700 pacotes que estendem o R. Diversas

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



funcionalidades são então acrescentadas, indo desde métodos específicos para análise estatística, até pacotes para a criação de interfaces gráficas. 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, ou FORTRAN (HENNING *et al.*, 2007).

3 R PARA CONTROLE ESTATÍSTICO DE PROCESSO

O Controle Estatístico de Processo é uma das duas importantes aplicações de engenharia focalizada neste documento. Para este tipo de aplicação o R dispõe de três pacotes específicos. O primeiro deles, o pacote *qcc* (Quality Control Charts) foi desenvolvido por SCRULLA (2004), 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. Por ser um ambiente GNU, o R foi um candidato natural. 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). Engloba também alguns conjuntos de dados da literatura da área, o que é importante do ponto de vista didático (HENNING *et al.*, 2007).

Na Figura 1 pode ser visualizado um exemplo ilustrativo do pacote *qcc*.

O pacote SPC (Statistical Process Control), criado por KNOTH (2007), enfoca outro aspecto: calcula ARLs (*average run lengths*) e limites para gráficos de controle CUSUM e EWMA. O mais recente dos pacotes é o *qAnalyst*, elaborado por SPANO (2008), que constrói gráficos de controle para variáveis e atributos, seguindo além da análise da capacidade de processos para distribuições normais e não normais.

Tratando-se do ambiente acadêmico, cabe destacar que mediante a utilização do R, o aluno pode, portanto resolver exercícios complexos, utilizando os pacotes e funções existentes, podendo se dedicar mais na interpretação dos resultados.

É também é possível criar seus próprios programas, para gráficos não contemplados pelos pacotes ou para assimilar melhor os conceitos que aprendeu. Pode também simular diversas situações, por exemplo, calcular o respectivo ARL. Além do próprio conteúdo trabalhado em aula, esses pacotes e funções permitem auxiliar nas atividades de pesquisa além de abrir novas frentes nesta.

ALVES *et al.* (2008) propõem o desenvolvimento de gráficos multivariados MCUSUM e MEWMA no R como uma ferramenta alternativa para análise estatística de processos multivariados. Estes gráficos vêm se destacando em diversos setores, tais como o metal-mecânico, automotivo, agrícola, de alimentos, entre outros.

A Figura 2 contém um exemplo da elaboração de uma rotina para um gráfico multivariado. Ressalta-se aqui a importância deste recurso computacional livre e de código aberto, na elaboração de funções ainda não disponíveis em alguns pacotes comerciais.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br

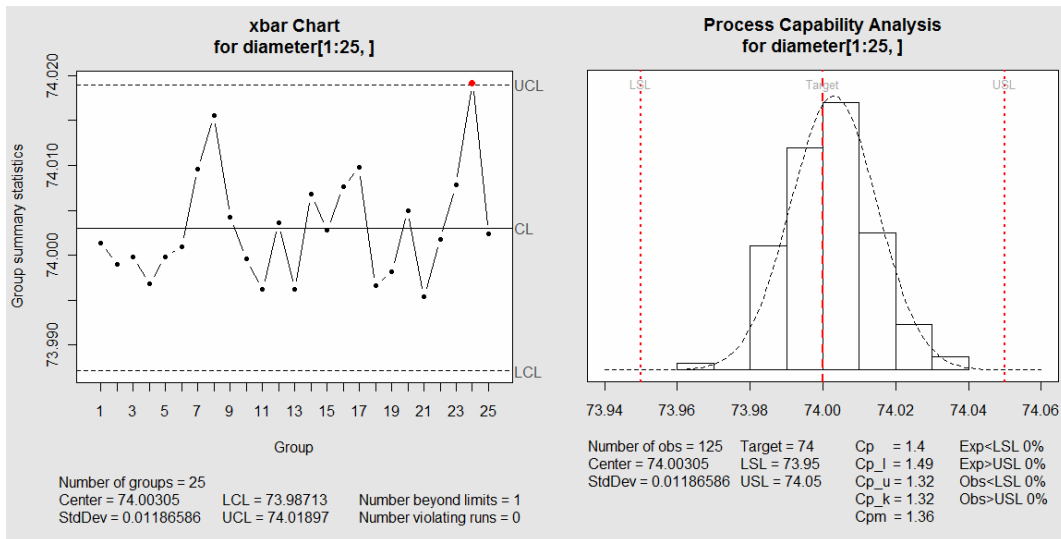


Figura 1: Exemplo de gráfico de controle e Análise de Capacidade de Processo com o pacote qcc. Fonte: R

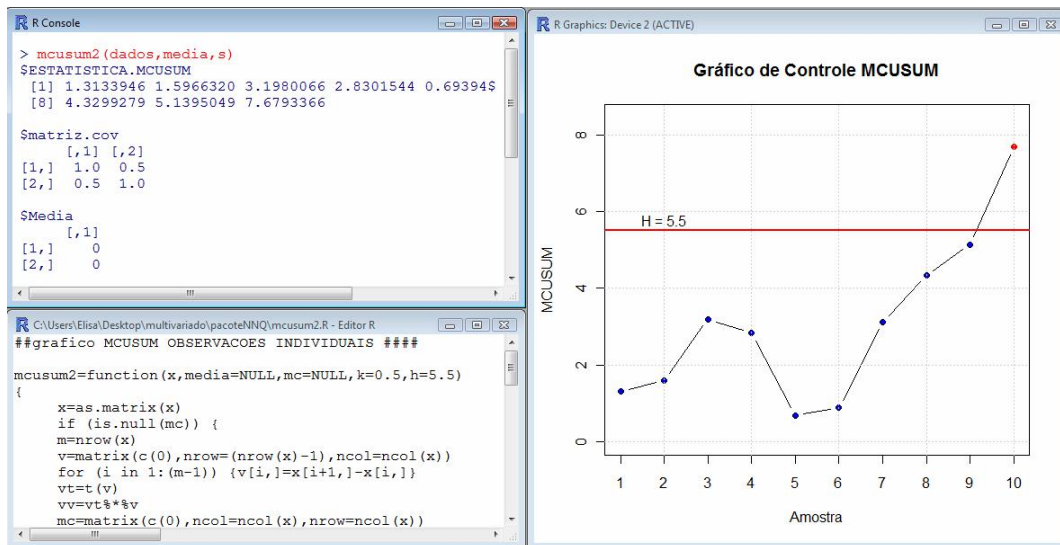


Figura 2: Exemplo de rotina desenvolvida no R para um gráfico de controle multivariado. Fonte: R

Uma dificuldade a ser pontuada é a interface de trabalho. O Rgui, que vem com a instalação usual do R, onde os comandos necessitam de digitação, pode encontrar resistência nos acadêmicos. Estes estão acostumados com planilhas e pacotes com interfaces gráficas.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



Assim, carregar dados e salvar resultados e arquivos podem implicar num trabalho extra para o mesmo. Por outro lado, o uso do editor de texto, simplifica as ações, no sentido que cada “exercício” ou atividade, pode ser armazenada na forma de uma rotina, podendo ser repetida quando necessário. Várias interfaces gráficas foram desenvolvidas para tornar o ambiente de trabalho do R “mais amigável”. Cita-se aqui o Rcmdr (FOX, 2005), inicialmente voltado para estatística básica. Um plug-in, para carregar as funções do qcc, nesta interface foi desenvolvido por CORTES (2009). Assim pode-se trabalhar num ambiente similar às planilhas e alguns pacotes comerciais mais populares neste meio.

Para gerar documentos o R tem funções que permitem trabalhar com o LaTeX, editor comum na área de matemática e estatística. Ressalta-se que é sempre possível salvar resultados numéricos na forma de planilhas eletrônicas e históricos em formato texto. O gráficos gerados podem ser salvos em diversos formatos.

4 MÉTODOS DE PREVISÃO EM ENGENHARIA

A análise de séries temporais é uma das importantes aplicações nas diversas áreas da engenharia. É importante, por exemplo, na engenharia de produção para realizar previsão de demanda e, na engenharia sanitária e ambiental na previsão de vazão de rios e precipitação, entre muitas outras variáveis. Segundo Souza & Camargo (2004) para realizar previsões de valores futuros, alguns objetivos são básicos: modelagem matemática do fenômeno, obtenção de conclusões e avaliação do modelo em termos de precisão.

O R dispõe de várias funções e pacotes para previsão de séries temporais. Inicialmente focalizamos o pacote *forecasting* desenvolvido por Hyndmann & Khandakar (2008), voltado à modelos de previsão. Este pacote possui procedimentos automáticos para escolha de modelos ARIMA e de suavização exponencial, entre outros.

A suavização exponencial é feita baseado nos modelos de classificação descritos em SOUZA, G. P. *et al.* (2008). O modelo é escolhido automaticamente, caso não haja nenhuma especificação. Para determinar os parâmetros de suavização, quatro critérios de otimização são disponibilizados. O default do programa considera o critério de máxima verossimilhança. A seleção dos modelos é efetuada com a metodologia stepwise utilizando os critérios de informação de Akaike (AIC), AICc e/ou Bayesiano (BIC). Além disso, o R retorna algumas medidas dos erros de previsão dentro e fora da amostra (SOUZA, G. P. *et al.*, 2008). A seleção de modelos ARIMA é feita de modo similar, com o programa atribuindo valores para os parâmetros das componentes do modelo, além de efetuar testes para verificar a estacionariedade e sazonalidade da série. A Figura 3 traz dois exemplos contendo séries com previsões pontuais e intervalares para os dois métodos anteriormente citados.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br

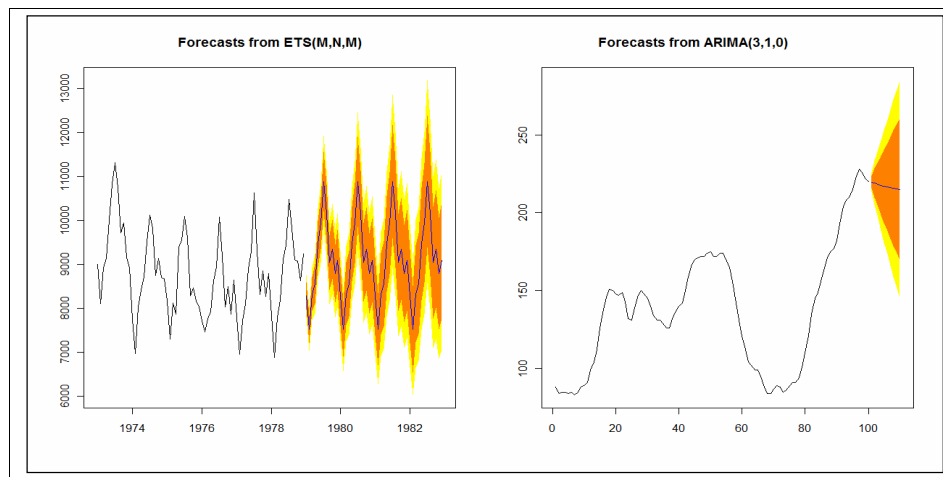


Figura 3: Exemplos do pacote forecast para previsão. Fonte: R

Os modelos de regressão múltipla, para Hanke (2001), tratam de problemas que envolvem o uso de mais de uma variável independente na previsão de uma variável dependente e alguns dos conceitos utilizados são os da análise de regressão simples. Neste caso, alguma teoria subjacente, envolvendo estas outras variáveis, pode explicar o comportamento da série. O problema de se trabalhar com modelos de regressão em dados de séries temporais é que raramente os erros são independentes ou não correlacionados, condição necessária para uma boa adequação do ajuste. Para a solução deste problema a regressão dinâmica é uma alternativa que permite a seleção de variáveis regressoras que possivelmente estão relacionadas com a variável dependente e que, portanto, explicam o comportamento de determinada série (ISHIKAWA & GIRARDI, 2006). Os métodos de regressão simples e múltipla estão na instalação padrão do R. Para a regressão dinâmica tem-se o pacote o dynlm (ZEILEIS, 2008).

Algumas deficiências citadas na seção anterior cabem também nesta. Complementa-se colocando que os procedimentos automáticos devem ser aplicados com olhar crítico, analisando-se a situação de acordo com a fundamentação teórica.

5 DIVULGAÇÃO

Para divulgar aplicações do R nos assuntos acima descritos, desenvolveu-se uma página (<http://www.qualimetria.ufsc.br>), onde material específico é disponibilizado. O objetivo principal é de auxiliar os estudantes e profissionais de engenharia a utilizar o software GNU R. Procurou-se pesquisar e escrever algo simples e didaticamente acessível a todos, com base na literatura da área. O acadêmico tem acesso a documentos que o auxiliam a fazer *download* do programa, contatos iniciais, estatística básica, controle estatístico de processos e métodos de previsão (Figura 4).

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



Figura 4: Página do Núcleo de Normalização e Qualimetria para o R.

6 COMPLEMENTAÇÕES

Alguns livros de estatística, voltados a cursos de engenharia, já possuem pacotes onde os conjuntos de dados e também exercícios estão disponíveis. Citam-se alguns mais conhecidos. É o caso do livro “Probabilidade e Estatística para Engenharia e Ciências”, de DEVORE (2006), em três pacotes específicos, conforme o número da edição. Os livros “Controle Estatístico de Qualidade” de MONTGOMERY (2004), e “Estatística Aplicada e Probabilidade para Engenheiros” de MONTGOMERY & RUNGER (2003), tem diversos conjuntos de dados utilizados nestes disponíveis no R. O livro de SÁ (2007), “Applied Statistics Using SPSS, STATISTICA, MATLAB and R”, aplica quatro recursos computacionais diferentes. Todos os conjuntos de dados da obra “Forecasting – Methods and Applications”, de MAKRIDAKIS *et al.* (2002), por exemplo, estão nos pacotes desenvolvidos pelo autor, que inclusive traz inúmeros exemplos.

As aplicações para o ensino de engenharia são inúmeras e aqui apenas algumas foram citadas. Um bom desafio seria incentivar os acadêmicos a explorá-las. Há pacotes que permitem executar funções do MATLAB® no R. As diversas funções compreendem também métodos numéricos para resolução de sistemas lineares e não lineares, integração numérica simples e múltipla, resolução de equações diferenciais, entre outros.

7 CONCLUSÕES E CONSIDERAÇÕES FINAIS

O ambiente GNU R é uma importante ferramenta para o ensino de métodos estatísticos nos cursos de engenharia. A sua divulgação e aplicação neste meio pode ser ampliada. Para o docente são inúmeras as vantagens. Como principal ponto positivo está o fato de ser um ambiente de código aberto e flexível, facilitando a utilização no ambiente acadêmico.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



Algumas deficiências são evidentes e, motivaram este trabalho. Há muito material no R, mas predominantemente em inglês. A interface padrão pode não ser muito atrativa aos acadêmicos acostumados a trabalhar com planilhas. Diversas interfaces gráficas foram desenvolvidas por pesquisadores. A escolha e/ou adaptação de uma interface recai num dos grandes problemas atuais: o excesso de opções. Mas, ao ser olhado por outro ângulo, isto é um ponto positivo: proporciona o acadêmico a enfrentar desafios diante de constantes mudanças e, sobretudo realizar várias escolhas para aperfeiçoar suas atividades.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, C. C.; HENNING, E.; SAMOHYL, R. W. O desenvolvimento de gráficos de controle MCUSUM e MEWMA em ambiente R como um procedimento alternativo para análise estatística de processos multivariados. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 27., Rio de Janeiro. **Anais...** Rio de Janeiro, RJ, Brasil, 2008.

BEASLEY, C. R. **Bioestatística usando R**. Universidade Federal do Pará. 2004. Disponível em: <<http://www.r-project.org/other-docs.html>> Acesso em: 01 set. 2005.

CORTES, R., **RcmdrPlugin.qcc**: Rcmdr qcc Plug-In. R package version 1.0-1, disponível em: <<http://cran.r-project.org/web/packages/RcmdrPlugin.qcc>> Acesso em: 17 mar. 2009.

DEVORE, J.L. **Probabilidade e Estatística para Engenharia e Ciências**. São Paulo: THOMSON, 2006.

FOX, J. The R Commander: A Basic-Statistics Graphical User Interface to R. **Journal of Statistical Software**, Vol. 14, 9, 2005. Disponível em: <<http://www.jstatsoft.org/>> Acesso em: 10 out. 2007.

GIRARDI, T. R.; ISHIKAWA, G. Previsão de curto prazo de carga de energia elétrica do estado de Santa Catarina. In: SIMPEP - SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 13. Bauru. **Anais: ...** Bauru: UNESP, 2006. Disponível em: http://www.simpep.feb.unesp.br/anais_simpep_aux.php?e=13 Acesso em: 04 mai. 2008.

HANKE, J. WICHERN, D. & REITSCH, A. **Business Forecasting**. 7ª Edição. New Jersey: Prentice Hall, 2000.

HENNING, E.; ALVES, C. C.; VIEIRA, V. O ambiente R como uma proposta de renovação para aprendizagem e monitoramento de processos em Controle Estatístico de Qualidade. In: SIMPEP - SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 14. Bauru. **Anais: ...** Bauru: UNESP, 2007. Disponível em: < http://www.simpep.feb.unesp.br/anais_simpep.php?e=1> Acesso em: 04 mai. 2009.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



HYNDMANN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, 26(3), 2008. Disponível em: <<http://www.jstatsoft.org/>> Acesso: em 20 jan. 2009.

KNOTH S. **spc: Statistical Process Control**. R package version 0.21, 2007. Disponível em: <<http://www.R-project.org>> Acesso em: 02 fev. 2009

MAKRIDAKIS S.; WHEELWRIGHT, S.T; HYNDMAN, R.J.; **Forecasting Methods and Applications**. 3th. Edition. John Wiley & Sons, 2002.

MONTGOMERY, D. C. **Introdução ao Controle Estatístico de Qualidade**. 4ª Edição. Rio de Janeiro: LTC, 2004.

MONTGOMERY, D. C; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 2ª Edição. Rio de Janeiro: LTC, 2003.

R DEVELOPMENT CORE TEAM (2008). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org>> Acesso em: 02 fev. 2009.

SÁ, J. M. **Applied statistics using SPSS, STATISTICA, MATLAB and R**. New York: Springer, 2007.

SCRUCCA, L. qcc: an R package for quality control charting and statistical process control. **R News** 4/1, 11-17, 2004. Disponível em: <<http://www.R-project.org>> Acesso em: 02 out. 2005.

SILVA, P. **Statistical education for doing Statistics professionally: some challenges and the road ahead**. Disponível em: <<http://www.maths.otago.ac.nz/icots7/data/PedroSilvaPlenaryICOTS.pdf>> Acesso em: 21 ago. 2006.

SOUZA, G. P., MIRANDA, R., SAMOHYL, R. W. **Métodos Simplificados em Previsão Empresarial**. Rio de Janeiro: Editora Ciência Moderna, 2008

SOUZA, R.C. & CAMARGO, M.E. **Análise e previsão de séries temporais: os modelos ARIMA**. Gráfica e Editora Regional, 2ª edição, Rio de Janeiro, 2004.

SPANO A. **qAnalyst: Variables and Attributes Control Charts**. R package version 0.5.4. Disponível em: <<http://www.quantide.com>> Acesso em: 10 out. 2008.

ZEILEIS, A. **dynlm: Dynamic Linear Regression**. R package version 0.2-1. Disponível em: <<http://www.r-project.org>> Acesso em: 09 nov. 2008.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br



A SHORT VIEW OF THE R ENVIRONMENT FOR ENGINEERING EDUCATION

***Abstract:** This short document presents and discusses some computational packages and applications from the GNU R for engineering education. The aim here is to increase the use of this programming tool. Applications are discussed in Statistical Quality Control and Forecasting Methods in Engineering. As a complement we present a web page under construction.*

Key-words: GNU R, Statistical Process Control, Prediction Methods in Engineering, Education.

Secretaria Executiva: Factos Eventos.

Rua Ernesto de Paula Santos 1368, salas 603/604. Boa Viagem Recife - PE CEP: 51021-330

PABX:(81) 3463 0871

E-mail: cobenge2009@factos.com.br