

Uso do *software* R no tratamento estatístico de dados na Engenharia.

Dominique Santos Ferreira

Raquel Cymrot

Universidade Presbiteriana Mackenzie
dominique.ferreira@mackenzista.com.br

Universidade Presbiteriana Mackenzie
raquel.cymrot@mackenzie.com.br

Resumo—O uso de ferramentas computacionais para auxiliar a aplicação de técnicas estatísticas é altamente recomendado, visto que possibilita que análises gráficas e métodos numéricos que necessitariam maior esforço, como no caso de uma grande quantidade de dados, de cálculos laboriosos ou de utilização de computação intensiva, sejam viabilizados corretamente e em tempo hábil. Contudo escolher adequadamente um *software* estatístico pode não ser tarefa fácil. Aplicativos estatísticos, em geral, são substancialmente caros ou de difícil compreensão. O uso de planilhas eletrônicas, como Microsoft Excel, na aplicação de técnicas estatísticas causa controvérsias, no que se refere à confiabilidade das saídas computacionais. Este artigo tem como objetivo apresentar o *software* R como uma solução viável de *software* livre na realização de análises estatísticas por estudantes e profissionais de Engenharia Elétrica, a partir de material didático embasado no ensino da Estatística aos cursos de Engenharia e também em pesquisas e teses da área.

Palavras chave—Engenharia, Estatística, *software* estatístico, tutorial.

Abstract—The use of computational tools to assist the application of statistical techniques is highly recommended, since can possibly statistical analysis and techniques, whereas it makes it possible that graphical analyses and numerical methods which would need much effort, such as a large amount of data, tiresome calculations or the use of intensive computation processing, can be made correctly and in a timely manner. However the correct choice software for a statistical application might not is an easy task. Statistical software is, in general, very expensive or hard to figure out. The use of spreadsheets, like Microsoft Excel, is disputed in regard of how reliable its computational outputs obtained from statistical techniques. This paper aims to present the R project as viable solution tool open software for statistical analysis by Electric Engineering students and professionals, based on courseware from Statistics classes of Engineering and field tests and researches.

Key words—Engineering, Statistics, statistical software, tutorial.

I. INTRODUÇÃO

A palavra Estatística, em inglês *Statistics* derivado do vocábulo *State*, referia-se em seu surgimento a coleção de

eventos de interesse do Estado. Ao ano de 1849, Quetelet elucidava o papel da Estatística como o de compor uma representação fiel do Estado numa determinada época. Já no século XVIII, tornou-se uma taquigrafia para “ciências descritivas do Estado” e por volta de 1830 o termo passou a ser sinônimo de “ciência numérica” da sociedade [1]. Atualmente o campo da estatística concatena métodos para organizar e tratar os dados para delinear conclusões a partir das informações que estes transmitem [2]. Ou ainda segundo Ross Estatística define-se como “a arte de se aprender a partir dos dados” [1].

Nos últimos vinte anos as práticas estatísticas aliaram-se ao advento da Computação, e pode-se afirmar que seja esta uma das principais razões que justifiquem sua intensa aplicação em trabalhos científicos. É imprescindível a engenheiros e cientistas o uso de ferramentas computacionais voltadas a Estatística, sendo assim indispensável seu conhecimento para interpretar as informações obtidas [3]. O uso do computador tornou consideravelmente mais rápida a realização de operações matemáticas, aumentou sua precisão, como também diminuiu o tempo empregado. Dada a sua capacidade de executar automaticamente tarefas e cálculos repetitivos, o auxílio do computador é fator determinante no custo e tempo empenhado [4].

Há uma variedade de *softwares*, pacotes estatísticos ditos genéricos por ter como finalidade a realização de uma vasta gama de métodos estatísticos, amplamente utilizada no meio acadêmico, em indústrias, bancos, órgãos públicos etc., como por exemplo, o Minitab, S-PLUS, SPSS e SAS. E ainda aplicativos matemáticos que detêm implementações de análises estatísticas como o Mathematica, o Maple, o Gauss e o MATLAB [5].

Esta pesquisa propõe o uso do R como ferramenta computacional, aplicada ao estudo da Estatística, por meio da elaboração de um tutorial. Tal tutorial deve auxiliar o aluno, de modo que o *software* atue como agente facilitador na execução de algumas análises estatísticas presentes na formação e no atendimento as solicitações em contextos próprios da Engenharia, na qual fica evidente a indispensabilidade da utilização de *softwares* estatísticos a fim de favorecer e viabilizar a aplicação de métodos estatísticos.

O tutorial deve ser um guia sintético do uso do aplicativo na realização de algumas análises estatísticas, como por exemplo, a construção do diagrama de caixa, elaboração de cartas de controle, geração de números pseudoaleatórios conforme

Manuscrito recebido em 13 de março de 2012; revisado em 13 de março de 2012.

D. S. Ferreira. (dominique.ferreira@mackenzista.com.br) e Orientadora Prof.^a Ms. R. Cymrot (raquel.cymrot@mackenzie.com.br) vinculadas a Escola de Engenharia do Instituto Presbiteriano Mackenzie. Rua da Consolação, 896 (prédio 6) – São Paulo – SP – Brasil – 01302-907. www.mackenzie.com.br

distribuições de probabilidade e construção de intervalos de confiança por meio da técnica de Bootstrap, considerada de computação intensiva, e muito útil no que diz respeito à inferência estatística quando as distribuições dos parâmetros não são conhecidas.

II. O QUE É AMBIENTE R?

R é o aplicativo e também a linguagem de programação, no qual são baseadas as suas rotinas. Seu intuito é integrar ferramentas gráficas, de cálculo e de análise de dados, em um ambiente de desenvolvimento de programação [6]. R também pode ser definido como uma “GNU S”, isto é, uma versão livre do *software* S [7]. GNU refere-se ao projeto GNU que promove a utilização e elaboração de aplicativos e sistemas livres. De acordo com o projeto o termo livre, *free* originalmente, está relacionado com liberdade e não com preço, pode-se obter ou não gratuitamente *software* GNU. O projeto baseia-se na premissa de colaboração entre desenvolvedores, um exemplo conhecido de implementação feita pelo projeto GNU é o sistema operacional Linux [8].

R foi desenvolvido por Robert Gentleman e Ross Ihaka do Departamento de Estatística da Universidade de Auckland [7]. Os direitos autorais do R são avaliados segundo a GNU General Public License (GPL) – Licença Pública Geral, ou seja, o autor detém a patente e em contrapartida permite-se legalmente aos usuários do programa o direito de copiá-lo, distribuí-lo, ou modificá-lo a partir do código-fonte, sob os termos da licença [9]. O CRAN (Comprehensive R Archive Network) é o conjunto de *sites* responsável pela alocação e distribuição dos arquivos referentes ao R, desde extensões contribuídas a arquivos de instalação e documentação [10]. Desde 1997, o R Core Team é responsável pelas atualizações ocorridas até então [6], R é colaborativo e por isso já detém algumas milhares de aplicações sob o seu código.

R é constituído por uma janela principal chamada *R Console*. O termo *console*, do inglês “painel de controle”, caracteriza a janela na qual os comandos são inseridos e executados, ou diretamente ou com uso de arquivos *scripts*.

R é mais que simplesmente um sistema computacional para estatística é um conjunto de recursos para manipulação de dados, cálculo e exibição gráfica. Por isso “ambiente” caracteriza-o como um sistema flexível e de vasta aplicabilidade [6].

A. Um software estatístico para Engenharia.

A sintaxe do R é similar a linguagens de programação como C e C++ [7], em características como o manuseio e tratamento de objetos por meio de funções, uso de bibliotecas de funções pré-definidas – *packages*, e de programação orientada ao objeto (POO). E por ser uma linguagem de programação a possibilidade de desenvolvimento de novas aplicações sob a sua linguagem é deveras atrativa, isto é, se não há ainda desenvolve-se um algoritmo a partir da linguagem.

A programação orientada ao objeto permite maior coerência e rentabilidade em uma aplicação alterando o modo como os comandos são descritos e permitindo que a aplicação implementada sob a linguagem de programação detenha maior proximidade com as características do objeto [4].

Essas características são algumas vantagens do uso desse

aplicativo por estudantes e profissionais da Engenharia, pois a grade curricular possibilita o contato com essas linguagens ou com *softwares*, como o MATLAB, que têm escopo semelhante.

Em sua legislação, o CREA-SP (Conselho Regional de Engenharia, Arquitetura e Agronomia do Estado de São Paulo), no cumprimento de suas atribuições, classifica em dezoito as atividades referentes ao exercício profissional das diferentes ramificações existentes na Engenharia, Arquitetura e Agronomia, dentre elas as seguintes:

“Atividade 02 - Estudo, planejamento, projeto e especificação;”

“Atividade 08 - Ensino, pesquisa, análise, experimentação, ensaio e divulgação técnica; extensão;”

“Atividade 10 - Padronização, mensuração e controle de qualidade;” [11].

O âmbito dessas atividades está imbuído do propósito da Estatística, isso se evidencia na definição dada por Montgomery e Runger para Estatística, a ciência que lida com a coleta, apresentação, análise e uso dos dados para tomar decisões, resolver problemas e planejar produtos e processos. O profissional de Engenharia tem de estar apto a lançar mão dessas técnicas para diante das situações cotidianas da área ampliar quantitativamente e qualitativamente seus resultados. E para tanto usufruir de uma ferramenta computacional se faz indispensável.

B. Por que utilizar o R?

Em uma pesquisa na qual se estudava os fatores avaliados na escolha de *softwares* estatísticos, com o intuito de definir qual seria o *software* que melhor adequar-se-ia no campus de uma Universidade pública. Evidenciou que o fator custo-benefício é determinístico no optar por um programa ou outro [12]. Aplicativos estatísticos são caros, o que impossibilita por muitas vezes a sua compra, o que remete também a aquisição ilegal destes *softwares* por seus usuários obstruindo assim questões autorais.

O uso de planilhas eletrônicas é uma das alternativas à análise de dados e de ferramenta de apoio à decisão. O MS Excel consolida-se como o mais utilizado pelo mercado em diversas aplicações, por deter sofisticados recursos, interface amigável, e de alta integração com os demais aplicativos existentes, nos ambientes para o qual foi desenvolvido [13]. Em contrapartida, há ressalvas sobre a confiabilidade de suas saídas computacionais no que se refere ao uso de ferramentas estatísticas.

Uma avaliação dos resultados obtidos em cinco diferentes planilhas eletrônicas realizada por Almiron *et al.* salienta que o uso destes aplicativos implica na falta da implementação de análises estatísticas e por muitas vezes em problemas na precisão da estimação de parâmetros, na geração de números pseudoaleatórios e também na apresentação do resultado com o número correto de algarismos significativos [14]. Levine *et al.* afirma que “[...] (os usuários) não levaram necessariamente em consideração a precisão e a integralidade dos resultados estatísticos que o Excel produz”, e que embora o Microsoft Excel detenha características muito atrativas, questões relativas ao custo, maior familiaridade e interatividade são os

reais motivos de sua alta adesão [15].

A inserção de computadores e *softwares* estatísticos também faculta um importante benefício educacional, por fomentar métodos de simulação. Segundo Navidi [3] a simulação faz com que os princípios fundamentais da Estatística tornem-se vivos.

De acordo Cymrot aliar um aplicativo ao estudo da Estatística permite a interação entre mercado de trabalho e meio acadêmico, de modo que se forme um profissional inteirado do dever para com a sociedade. No entanto, o conhecimento e benefícios, adquiridos com a utilização de um aplicativo voltado à Estatística, só se dão plenamente ao ser capaz de discernir corretamente qual tratamento estatístico empregar, em função do contexto investigado [16].

O documento para “Declaração Mundial sobre Educação Superior no Século XXI: Visão e Ação” é o reflexo da preocupação de transformar o ensino superior numa ferramenta de via de acesso à educação e desenvolvimento. Onde podemos ressaltar que o curso e suas componentes curriculares devem exacerbar o compromisso ético do estudante e futuro profissional em trabalhar para o bem comum, abordada no artigo 1º, intensificar as relações entre o mundo de trabalho e o meio acadêmico, mencionada no artigo 7º, e por fim a oportunidade de agregar valor ao ensino por meio de novas tecnologias, citada no artigo 12º. [17]

III. EXEMPLOS DE APLICAÇÃO DO R.

O tutorial baseou-se na busca em base de dados como a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), bibliotecas digitais de teses e dissertações e em vasta literatura voltada ao ensino da Estatística à Engenharia. Visando suprir as necessidades por análises estatísticas indisponíveis em programas não específicos, como o Excel, como também valer-se de um *software* acessível aos alunos e profissionais da Engenharia, utilizando exemplos da área.

Abaixo constam explanações sobre análises estatísticas que compõem o tutorial, de forma a aludir à estrutura deste. Esta não é uma citação direta do conteúdo do tutorial, no entanto, o propósito ao dispor estas análises é demonstrar o funcionamento do R.

R é uma linguagem de programação atrelada a um ambiente de desenvolvimento próprio, por isso inicialmente pode ser desanimador lidar com uma interface gráfica não muito amigável, seus recursos são acionados, em suma, inserindo-se a partir do teclado os comandos necessários.

A. Regressão Linear Múltipla

No artigo “*Fatores que influenciam o desempenho de Receptores na TV Digital Terrestre na cidade de São Paulo*” um dos métodos utilizados para estudo dos fatores de degradação do sinal da TV Digital foi o da regressão linear múltipla [18]. A regressão múltipla ajusta um modelo linear entre uma variável de resposta e uma ou mais variáveis explicativas (regressoras).

O modelo de regressão linear múltipla requer uma abordagem matricial para sua resolução, se ao invés de efetuar os cálculos manualmente, utilizar-se uma abordagem computacional um maior número de análises poderá ser

realizado o que consequentemente propicia maior tempo de observação das características dos dados e melhor verificação dos critérios de adequação ao modelo [19].

$$y = X\beta + \varepsilon \quad (1)$$

Onde as variáveis são:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2)$$

A regressão linear múltipla relaciona linearmente um parâmetro a um ou mais variáveis pelo método dos mínimos quadrados. Abaixo vamos utilizar os dados da Tabela I para exemplificar um modelo de regressão múltipla, os dados são do exemplo 12-1 do livro “*Estatística Aplicada e Probabilidade para Engenheiros*” (cap. 12 pág. 267-271), onde duas variáveis independentes são usadas para explicar a resistência à tração de um fio colado obtido na fabricação de semicondutores [19].

TABELA I
RESISTÊNCIA À TRAÇÃO DO FIO COLADO.

Número da Observação	Resistência à Tração y	Comprimento do Fio x ₁	Altura da Garra x ₂
1	9,95	2	50
2	24,45	8	110
3	31,75	11	120
4	35,00	10	550
5	25,02	8	295
6	16,86	4	200
7	14,38	2	375
8	9,60	2	52
9	24,35	9	100
10	27,50	8	300
11	17,08	4	412
12	37,00	11	400
13	41,95	12	500
14	11,66	2	360
15	21,65	4	205
16	17,89	4	400
17	69,00	20	600
18	10,30	1	585
19	34,93	10	540
20	46,59	15	250
21	44,88	15	290
22	54,12	16	510
23	56,63	17	590
24	22,13	6	100
25	21,15	5	400

O modelo ajustado é:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (3)$$

β_0 é o intercepto ou intersecção do plano. β_1 e β_2 , chamados de coeficientes parciais de regressão, relacionam a resistência à tração respectivamente ao comprimento do fio – x_1 , e a altura da garra – x_2 , [19]. O modelo deve ater-se as seguintes condições:

(a) y não é variável aleatória;

(b) Os erros são normalmente distribuídos;

(c) Os erros devem ter média $\mu_e = 0$ e variância constante σ_e^2 .

Com o auxílio do R vamos identificar e testar o modelo proposto, para tanto foram usadas as versões 2.13.1 e 2.14.0 do R. Abaixo o código gerado a partir de um arquivo *script*. O código descrito a seguir pode ser compilado, bastando copiar os comandos no R *Console*, as linhas de comandos estão separadas por espaçamento tornando-as parecidas com parágrafos, a numeração vermelha entre parênteses representa a posição em que foi listado o comando, para usar o código no R desconsiderar a numeração, os comentários não forma considerados linhas de comando.

```
(1) # Comentário
    # Exemplo de Regressão Múltipla no R.
# Armazenando os dados em vetores.
y <- c(9.95, 24.45, 31.75, 35.00, 25.02, 16.86,
14.38, 9.60, 24.35, 27.50, 17.08, 37.00, 41.95,
11.66, 21.65, 17.89, 69.00, 10.30, 34.93, 46.59,
44.88, 54.12, 56.63, 22.13, 21.15)

x1 <- c(2, 8, 11, 10, 8, 4, 2, 2, 9, 8, 4, 11,
12, 2, 4, 4, 20, 1, 10, 15, 15, 16, 17, 6, 5)

x2 <- c(50, 110, 120, 550, 295, 200, 375, 52,
100, 300, 412, 400, 500, 360, 205, 400, 600, 585,
540, 250, 290, 510, 590, 100, 400)

# Lista os objetos existentes.
ls() # objects()

(5) # Resumo dos cinco números da Resistência
Tração.
summary(y); # fivenum()

# Boxplot da Resistência Tração
boxplot(y, notch=T, col=2);

# Matriz de gráficos de dispersão.
pairs(y~x1+x2, pch="@", panel=panel.smooth, labels=c
("Resistência", "Comprimento", "Altura"))

# Ajuste do modelo:
model <- lm(y~x1+x2) # y~x1+x2 a variável y é
explicada pelas variáveis x1 e x2

model # Exibe os resultados do modelo.

(10) summary(model) # Análise de regressão

# Análise de Variância
anova(model)

# Análise Residual
par(mfrow=c(2,2))

qqnorm(residuals(model), datax=T, pch=20, main="Gráfico
de Probabilidade Normal",
xlab=" ", ylab="Resíduos")

qqline(residuals(model), datax=T, col="red")

(15) plot(residuals(model), pch=8, type="b",
col="red", main="Resíduos x
Ordem", ylab="Resíduos", xlab="Ordem das
observações")

plot(fitted.values(model), residuals(model),
```

```
ylab="Resíduos", xlab="Valores Ajustados da
Resistência", pch=20, col=2, main="Resíduos x
Valores Ajustados")
```

```
abline(0,0,lty=3)
```

```
hist(residuals(model), prob=T,
breaks=seq(-4,6,by=0.5), ylim=c(0,0.35),
yaxp=c(0,0.35,5), xaxp=c(-4,6,10),
main="Histograma", ylab="Densidade",
xlab="Resíduos")
```

```
lines(density(residuals(model)))
```

```
(20) # Gráfico 3-D (y,x1,x2)
x1 <- seq(0,20,by=0.1) # Valores arbitrados para
o comprimento.
```

```
x2 <- seq(0,600,length=101) # Valores arbitrados
para a altura.
```

```
beta <- coef(model) # Coeficientes do Modelo
```

```
f <- outer(x1,x2,function(x1,x2){
beta[[1]]+beta[[2]]*x1+beta[[3]]*x2})
```

```
persp(x1,x2,f,theta=-45,phi=30,col=rainbow(100),
zlab="Resistência", xlab="Comprimento",
ylab="Altura")
```

```
(25) # O código acaba aqui.
```

Parece intuitivo que a função `boxplot()` gere como saída computacional um diagrama de caixa ou a função `summary()` summarize os dados, isto é, exponha algumas medidas descritivas a partir de objetos que representem os dados. No entanto, ao deparar-se com as funções `par()` ou `outer()` não fica claro o quê desempenham. Por isso familiarizar-se com uma linguagem de programação não é tarefa fácil, mas de grande valia. Abaixo vemos a saída da função `summary()` para os valores medidos da Resistência à Tração do fio Colado representada pelo vetor 'y', um resumo dos cinco números.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
9.60	17.08	24.45	29.03	37.00	69.00

As saídas do R são imediatas, ou melhor, interativas, ao se digitar um comando visualiza-se os resultados, a não ser quando redirecionados para outros arquivos ou guardadas sob a forma de objeto, como é o caso do objeto *model*, só vemos os resultados do modelo ajustado pela função `lm()` – *linear model*, ao inserir o identificador – nome do objeto, no R *Console* como na linha de comando número 9. E no caso do objeto *model* o redirecionamento citado deu-se a partir do operador de atribuição (`<-`), que é um operador de atribuição como o símbolo de igual em C e C++.

Todas as funções tem o que se chama de argumentos que definem o comportamento de suas saídas computacionais e os próprios dados de entrada são argumentos. Na função `boxplot()` o argumento “`notch=T`” habilita os entalhes vistos em cada lado no boxplot, esse recurso serve para melhor visualização na hora de comparar dois boxplots numa janela gráfica a partir da sobreposição dos entalhes se os entalhes se sobrepuserem há uma forte evidência de que as medianas não diferem.

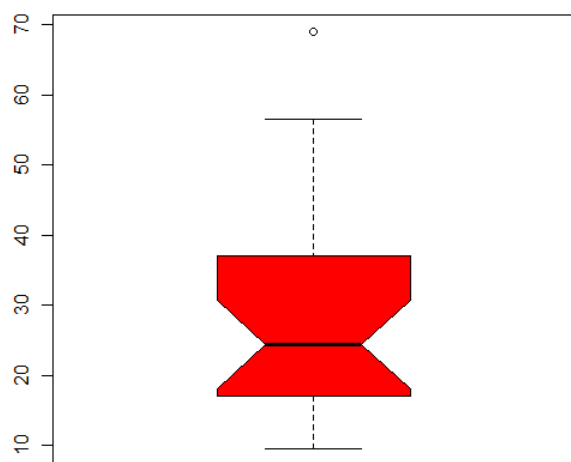


Fig. 1. Diagrama de Caixa (Boxplot) dos valores da Resistência à Tração do Fio Colado.

A partir do R obteve-se também saídas gráficas, o Boxplot da Resistência à Tração do Fio, na figura 1, as relações entre as variáveis do modelo pela matriz de dispersão, na figura 2, onde a proposta do modelo de regressão linear toma forma, o comprimento e a resistência parecem estar linearmente relacionados. Em relação à altura e a resistência não há como afirmar que sim ou não, os testes em função do modelo indicarão a sua validade.

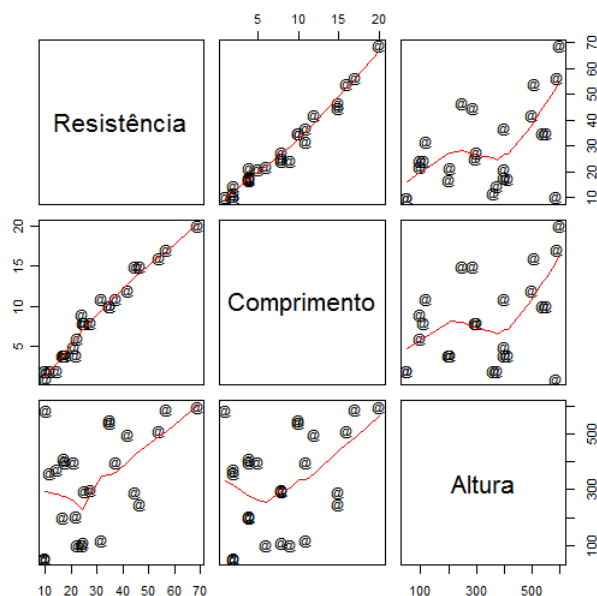


Fig. 2. Possíveis interações entre as variáveis do problema numa matriz de gráfico de dispersão, função pairs().

A saída da função `lm()` armazenada no objeto *model* calcula os coeficientes $\beta_0 = 2,26379$, $\beta_1 = 2,74427$ e $\beta_2 = 0,01253$. Seguida da saída da função `summary()` para o objeto *model*, é importante ressaltar novamente a flexibilidade dada pela programação orientada ao objeto (POO), a mesma função desempenha papéis diferentes quando do uso por objetos diferentes, na variável *y*, que é um vetor, foi realizado o resumo dos cinco número dos dados da resistência, já com o objeto *model* um resumo das informações do modelo ajustado.

```
Call:
lm(formula = y ~ x1 + x2)

Coefficients:
(Intercept)      x1      x2
  2.26379    2.74427    0.01253
```

```
Call:
lm(formula = y ~ x1 + x2)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.865 -1.542 -0.362  1.196  5.841
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.263791    1.060066   2.136  0.044099
x1           2.744270    0.093524  29.343 < 2e-16
x2           0.012528    0.002798   4.477  0.000188
---
```

```
Residual standard error: 2.288 on 22 degrees of freedom
```

```
Multiple R-squared:  0.9811,    Adjusted R-squared:  0.9794
```

```
F-statistic: 572.2 on 2 and 22 DF, p-value: < 2.2e-16
```

```
---
```

Na saída da função `summary()` para o objeto *model* são dispostas informações importantes como o resultado do teste T para os coeficientes da regressão e do teste de significância da regressão.

A análise residual verifica se o modelo obedece as seguintes suposições: (a) Os resíduos devem ser normalmente distribuídos; (b) A média dos resíduos deve ser zero; (c) A variância, σ^2 , deve ser igual a uma constante e (d) Os resíduos devem ser independentes [19]. De forma que o modelo só é válido a respeitando-se essas premissas.

Essas suposições podem ser verificadas graficamente no Gráfico de Probabilidade Normal pela regra do “lápis grosso”, na qual se for possível à identificação de uma reta que possa concatenar a maioria dos valores residuais a suposição é satisfeita, no Histograma dos resíduos também é possível verificar essa suposição – no caso do Histograma recomenda-se que haja várias medidas, isto é, que a amostra seja de grande tamanho. Em ambos os casos, o melhor a se fazer é a realização de um teste de aderência à distribuição normal, teste de normalidade, onde de fato a suposição seria testada o R oferece uma boa gama de testes de normalidade e de aderência a distribuições quaisquer, um *package* chamado *nortest*, criado por Juergen Gross, agrega ao R cinco diferentes testes de normalidade dentre eles o de Lilliefors e o de Anderson-Darling.

A suposição (d) se constata a partir do gráfico de resíduos versus a ordem de experimentação, onde os resíduos não devem seguir nenhum padrão e sim apresentar aleatoriedade. A hipótese de variância constante (c) pode ser observada no gráfico de resíduos versus valores ajustados, espera-se observar que não haja aumento da dispersão dos resíduos em função dos diferentes valores ajustados, no mesmo gráfico

observa-se se há simetria aproximada dos pontos em relação ao valor zero de modo que a suposição (b) seja cumprida.

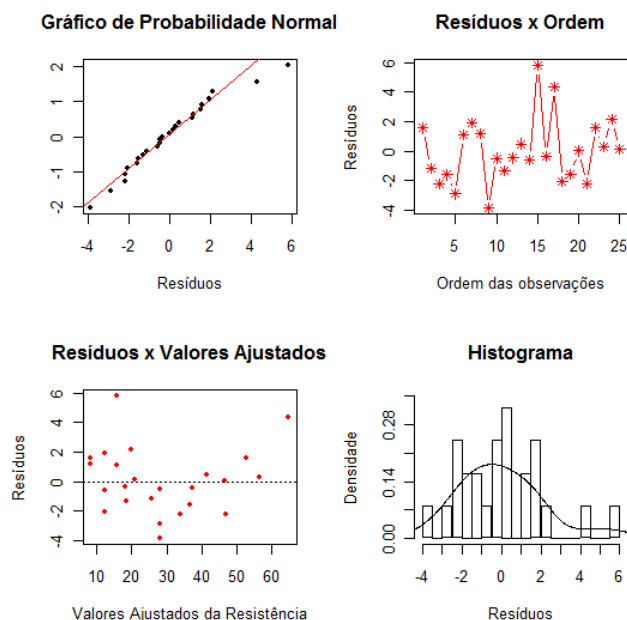


Fig. 3. Análise Residual do modelo para Resistência à Tração do Fio Colado.

Além disso, numa regressão estimam-se parâmetros como a variância do erro identificada na análise de variância efetuada na linha 11 do código. O resultado exposto pela função `anova()`, a variância do erro é igual a 5,2.

Analysis of Variance Table

```

Response: y
          Df Sum Sq Mean Sq  F value    Pr(>F)
x1          1  5885.9   5885.9  1124.293 < 2.2e-16
x2          1   104.9    104.9   20.041 0.0001883
Residuals  22    115.2      5.2

```

O gráfico da figura 4 é resultado da função `persp()` onde foram plotados valores esperados da resistência em função de valores de comprimento e da altura escolhidos de forma a não haver extrapolações do modelo ajustado.

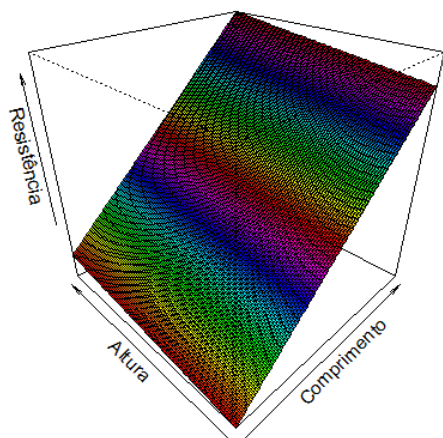


Fig. 4. Gráfico da equação proposta pelo modelo.

B. Gráficos de Controle

Gráficos ou Cartas de Controle são utilizadas no monitoramento de características de interesse a qualidade do processo

IV. CONCLUSÕES

O uso de *softwares* estatísticos é necessário, e deve ser amplamente difundido principalmente na área de Engenharia, em que profissionais e estudantes da área tem de estar aptos à análise de dados a fim de propor melhorias e de prover novas técnicas e tecnologias atendendo as necessidades atuais e vindouras da sociedade.

A análise estatística propicia a compreensão e julgamento dos mais diferentes fenômenos, com o uso de *softwares* e a implantação de laboratórios de computação voltados a Estatística o foco dos estudos pode concentrar-se no processo de desenvolvimento do aprendizado dos procedimentos estatísticos e não em como são realizadas operações matemáticas.

Em trabalhos e pesquisas acadêmicas, a verificação das características que os dados apresentam ou o estudo e ajuste de modelos propostos e ainda a inferência acerca de um parâmetro, podem ser elaboradas de forma mais eficiente a obter-se resultados em menor prazo.

Já no que tange a vida profissional a aplicação de Estatística é vinculada a ferramentas computacionais, na qual para muitos estudos e pesquisas faltam ferramentas de análise adequadas (preço, opções de ajuda ao usuário, qualidade das saídas computacionais), e de embasamento teórico necessário ao real aproveitamento dos dados.

Portanto o tutorial de um *software* como o R pode estimular a utilização de métodos estatísticos por parte dos profissionais e estudantes de Engenharia.

AGRADECIMENTO

Ao colega de estudos Lenin Umeji Abe por auxiliar na elaboração do presente artigo, sinceros agradecimentos.

A Prof.^a Ms. Raquel Cymrot orientadora sem a qual não se teria iniciado esse trabalho, muito obrigada.

REFERÊNCIAS

- [1] S. M. ROSS. *Introduction to Probability and Statistics for Engineers and Scientists*. 3. ed. San Diego: Elsevier Academic Press, 2004.
- [2] J. L. DEVORE. *Probabilidade e Estatística para Engenharia e Ciências*. 6. ed. São Paulo: Pioneira Thomson Learning, 2011.
- [3] W. C. NAVIDI. *Statistics for Engineers and Scientists*. Boston: McGraw-Hill, 2006.
- [4] S. V. D. PAMBOUKIAN, L. C. ZAMBONI e E. A. R. BARROS, *Aplicações científicas em C++: da programação estruturada à programação orientada a objetos*. / Sergio V. D. Pamboukian, Linconl C. Zamboni, Edson A. R. Barros. São Paulo: Páginas & Letras Editora e Gráfica, 2010.
- [5] P. A. MORETTIN e W. O. BUSSAB, *Estatística Básica* / Pedro A. Morettin, Wilton O. Bussab. 6. ed. São Paulo: Saraiva, 2010.
- [6] W. N. VENABLES, D. M. SMITH e o R DEVELOPMENT CORE TEAM. (2011, Agosto). *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Version 2.13.1*. URL: <http://www.r-project.org/>
- [7] R CORE TEAM (2001, Janeiro). *What's R? R News: The Newsletter of the R Project* [Online], vol. 1, no. 1, pp. 2-3. URL: http://cran.r-project.org/doc/Rnews/Rnews_2001-1.pdf

- [8] FREE SOFTWARE FOUNDATION. (2011, Dezembro). *Introdução ao Projeto GNU*. URL: <http://www.gnu.org/gnu/gnu-history.html>
- [9] GNU General Public License. FREE SOFTWARE FOUNDATION. Versão 2: Junho, 1991.
- [10] K. HORNIK. *The R FAQ*. URL: <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>
- [11] Resolução N° 218, de 29 de Junho de 1973. CONSELHO REGIONAL DE ENGENHARIA, ARQUITETURA E AGRONOMIA DO ESTADO DE SÃO PAULO CREA-SP. *Manual de Fiscalização e Legislação – São Paulo*. URL: <http://www.creasp.org.br/arquivos/publicacoes/legislacao.pdf>
- [12] J. F. da S. COSTA, M. G. CORREIA e L. T. T. de SOUZA, *Auxílio à Decisão utilizando o Método AHP - Análise Competitiva dos Softwares Estatísticos*. In: XXX ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO 2010. São Carlos, 2010. URL: http://www.abepro.org.br/.../enegep2010_TN_STO_113_739_16433.pdf
- [13] A. M. MIGLIOLI; L. H. OSTANEL; W. TACHIBANA. *Planilhas eletrônicas como ferramentas para apoio à decisão e geração de conhecimento na pequena empresa*. In: XXIV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO 2004, Anais..., Florianópolis, 2004. URL: http://www.abepro.org.br/biblioteca/ENEGEp2004_Enegep0902_1706.pdf
- [14] M. G. ALMIRON, B. LOPES, A. L. C. OLIVEIRA, A. C. MEDEIROS E A. C. FRERY. (2010, Abril). On the Numerical Accuracy of Spreadsheets. *Journal of Statistical Software*. [Online], vol. 34, issue 4. URL: <http://www.jstatsoft.org/v34/i04/paper>
- [15] D. M. LEVINE, D. F. STEPHAN, T. C. KREHBIEL e M. L. BERENSON. *Estatística: Teoria e Aplicações / David M. Levine... [et al.]*. 5. ed. Rio de Janeiro: LTC, 2008.
- [16] R. CYMROT. *A Utilização do Programa Excel no Aprendizado de Técnicas Estatísticas Utilizadas na Metodologia Seis Sigma*. In: WORLD CONGRESS ON COMPUTER SCIENCE, ENGINEERING AND TECHNOLOGY EDUCATION 2006 (WCCSETE), São Paulo, 2006. URL: <http://meusite.mackenzie.com.br/raquelc/WCCSETE2006Excel.pdf>
- [17] UNESCO. *Declaração Mundial sobre Educação Superior no Século XXI: Visão e Ação – 1998*. Biblioteca Virtual de Direitos Humanos da USP. URL: <http://www.direitoshumanos.usp.br/index.php/Direito-a-Educacao/declaracao-mundial-sobre-educacao-superior-no-seculo-xxi-visao-e-acao.html>
- [18] P. G. ESPERANTE, R. CYMROT, C. AKAMINE, F. YAMADA, R. E. MOTOYAMA e F. RAIA. (2010). *Fatores que influenciam o desempenho de Receptores na TV Digital Terrestre na cidade de São Paulo*. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE ENGENHARIA DE TELEVISÃO – 2010. Revista de Radiodifusão - SET [Online], vol. 4, n. 04. URL: <http://www.set.com.br/revistaeletronica/radiodifusao/index.php/revistaderadiodifusao/article/viewFile/45/55>
- [19] D. C. MONTGOMERY e G. C. RUNGER. *Estatística Aplicada e Probabilidade para Engenheiros*. 4. ed. Rio de Janeiro: LTC, 2009.
- [20] J. GROSS. (2006, Fevereiro). *nortest: Tests for Normality*. R package, versão 1.0.