

SOFTWARE LIVRE E SUAS APLICAÇÕES EM ESTATÍSTICA

Priscila B. Reis, Hítalo D. M. Chagas, Gabriel N. Almeida

UFMG

pribreis@gmail.com, hitalodias@yahoo.com.br, gabufmg@hotmail.com

RESUMO

Neste artigo descrevemos como o R é um ambiente importante, visto que, outros pacotes estatísticos equivalentes possuem um preço bastante elevado e, nem sempre existem versões estudantes disponíveis. Com isso o R é cada vez mais utilizado no meio acadêmico e não é difícil encontrar pesquisas nas áreas equivalentes à Estatística que utilizem este ambiente.

PALAVRAS-CHAVE: Ambiente R, software livre, pacotes estatísticos, Estatística.

1. INTRODUÇÃO

Entendemos a Estatística como um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados[1]. Como a maioria das pesquisas envolve uma grande quantidade de dados, o uso de pacotes estatísticos para a análise de dados é imprescindível. Contudo, estes pacotes apresentam um custo elevado e nem sempre possuem uma versão de estudante. Dentre os softwares de domínio público, livres, que podem ser utilizados para análise de dados em geral, encontra-se o Ambiente R, que apresenta código fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos por qualquer usuário a qualquer momento. Além do que, o R conta com um grande número de colaboradores das mais diversas áreas do conhecimento. O R é portanto, uma importante ferramenta na análise e na manipulação de dados, com testes paramétricos e não paramétricos, modelagem linear e não linear, análise de séries temporais, análise de sobrevivência, simulação e estatística espacial, entre outros, além de apresentar facilidade na elaboração de diversos tipos de gráficos, no qual o usuário tem pleno controle sobre o gráfico criado[2].

2. ALGUMAS APLICAÇÕES EM ESTATÍSTICA

2.1 Estatísticas Descritivas

Utilizamos a Estatística Descritiva para sumarizar as informações disponíveis sobre o comportamento de uma determinada variável.

```
> Idade <- c(17, 18, 19, 20, 21, 22, 23, 25, 24)
> mean(Idade) # cálculo da média
[1] 21
> median(Idade) # cálculo da mediana
[1] 21
> var(Idade) # cálculo da variância
[1] 7.5
> sd(Idade) # cálculo do desvio-padrão
[1] 2.738613
> max(Idade) # valor máximo
[1] 25
> min(Idade) # valor mínimo
[1] 17
```

Figura 2: Exemplos de comandos e saída de Estatísticas Descritivas no R

2.2 Gráficos

Os gráficos são muito utilizados na análise de dados são. Com eles podemos inclusive verificar a distribuição das variáveis. Dois gráficos bastante utilizados para essa finalidade são o Histograma e o Box-Plot.

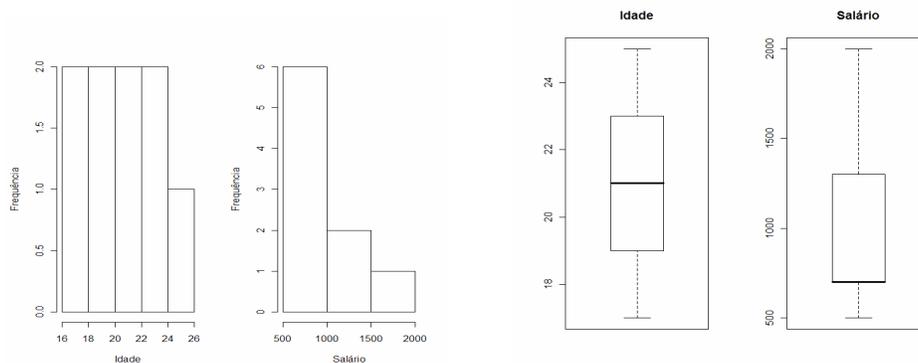


Figura 3: Histogramas e Box-Plot das variáveis Idade e Salário

2.3 Regressão

O principal objetivo da análise de regressão é verificar se existe uma relação entre duas variáveis quantitativas e quantificar essa relação tentando formular uma

relação direta entre uma ou mais variáveis independentes e seu efeito na variável dependente[3].

```

> Salário<-c(500,500,700,700,700,800,1300,2000,1500)
> #Fazendo a regressão linear simples.
> reg1<-lm(Salário~Idade)
> predict(reg1) #obtendo os valores estimados
      1          2          3          4          5          6          7          8
280.0000 451.6667 623.3333 795.0000 966.6667 1138.3333 1310.0000 1653.3333
                                           14.816.667
> coefficients(reg1) #obtendo as estimativas dos coeficientes
(Intercept)      Idade
-2638.3333      171.6667
> summary(reg1) #obtendo alguns resultados úteis: teste t e ANOVA
Call:
lm(formula = Salário ~ Idade)

Residuals:
    Min       1Q   Median       3Q      Max
-338.33  -95.00   18.33   76.67  346.67

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2638.33      629.55  -4.191 0.004082 **
Idade         171.67       29.75   5.769 0.000685 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 230.5 on 7 degrees of freedom
Multiple R-squared:  0.8262,    Adjusted R-squared:  0.8014
F-statistic: 33.29 on 1 and 7 DF,  p-value: 0.0006846

> anova(reg1)
Analysis of Variance Table

Response: Salário
      Df Sum Sq Mean Sq F value    Pr(>F)
Idade   1 1768167 1768167  33.287 0.0006846 ***
Residuals 7  371833   53119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #Para construir o gráfico
> plot(Idade,Salário) #construindo o diagrama de dispersão
> abline(reg1) #para obter a reta
> pred<-predict(reg1) #usando os valores estimados
> for(i in 1:9) lines(c(Idade[i],Idade[i]),c(Salário[i],pred[i]))

```

Figura 4: Exemplos de comandos e saída de Regressão no R

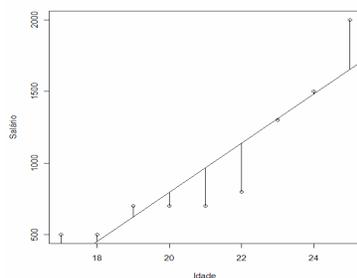


Figura 5: Gráfico de Regressão Linear no R

3. CONCLUSÕES E OBSERVAÇÕES FINAIS

O ambiente R é de grande importância no meio acadêmico por ser um Software Livre, distribuído sob a Licença Pública Geral e pode ser livremente copiado e distribuído entre usuários, bem como pode ser instalado em diversos computadores livremente. Isso contrasta com pacotes comerciais que têm licenças altamente restritivas e não permitem que sejam feitas cópias ou que seja instalado em mais de um computador sem a devida licença. O fato do código ser aberto possibilita que as falhas possam ser detectadas e corrigidas rapidamente e atualizações possam ser disponibilizadas rapidamente[4].

Se por um lado o R apresenta grande flexibilidade para usuários avançados, por outro, pode constituir uma desvantagem para usuários acostumados com ambientes gráficos mais interativos. Outra desvantagem do R é que como boa parte dos Softwares Livres não possuem tutoriais detalhados.

4. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Magalhães, M. N. , Lima, A. C. P. , *Noções de Probabilidade e Estatística*, EDUSP, São Paulo, 2007.
- [2]<http://www.ecologia.ufrgs.br/~adrimelo/lm/apostilas/Mini_tutorial_R.pdf >Acesso em: 20 Maio 2011, 18:50.
- [3]<<http://www2.ufersa.edu.br/portal/view/uploads/setores/137/Apostilas%20e%20Tutoriais%20-%20R%20Project/Apostila%20R%20-%20GenMelhor.pdf>>Acesso em: 20 Maio 2011, 22:49.
- [4]<<http://cran.r-project.org/doc/contrib/Beasley-BioestatisticaUsandoR.pdf>>Acesso em: 21 Maio 2011, 12:55.