Universidade Federal do Ceará

Programa de Pós-Graduação em Sistemática, Uso e Conservação da Biodiversidade (PPGSIS)

Disciplina: Análise de dados em R - uma abordagem prática para analise, visualização e modelagem

Prof. Samuel Carleial Fernandes

Aluno: Jhones de Lima Vieira, 474681, doutorando em Ciências Marinhas Tropicais - Labomar

**Atividade 1: Gráfico em R**

- Dois gráficos feitos em R (com código e figura lado a lado)

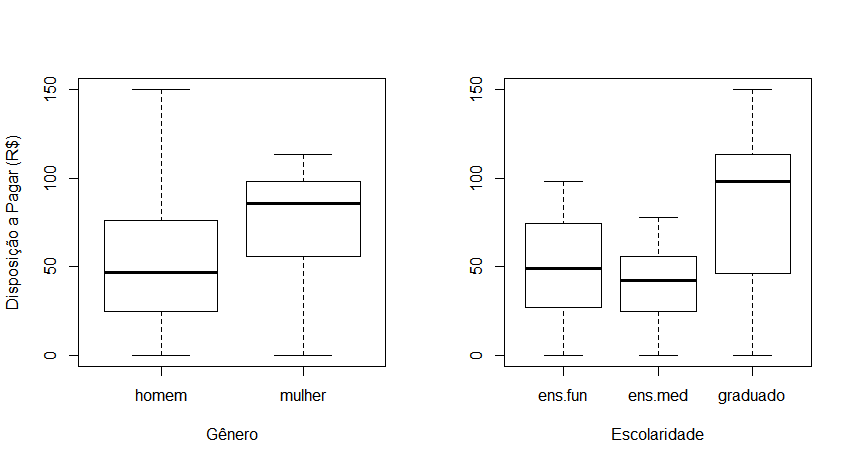


Fig 1. Usuários do sexo feminino e com nível superior completo estão dispostos a contribuir mais com políticas públicas de preservação ambiental.

- Código em R

|  |
| --- |
| # Elaboração de 2 gráficos no R  par(mfrow = c(1,2))  boxplot(DAP ~ sexo, data=valoracao\_ambiental,  xlab = "Gênero",  ylab = "Disposição a Pagar (R$)")  boxplot(DAP ~ grad, data = valoracao\_ambiental,  xlab = "Escolaridade",  ylab = NULL) |

- Pequena explicação do motivo/foco de cada gráfico

Os gráficos foram elaborados para observar se o gênero e a escolaridade contribuíam para aumentar a disposição a pagar de 30 produtores de camarão pela criação de um fundo de investimento para conservação da bacia hidrográfica do Rio Jaguaribe. Em relação ao gênero, o primeiro gráfico mostrou que a mediana de DAP das mulheres foi superior à dos homens. E em relação à escolaridade, o segundo gráfico indicou/sugere que os produtores com nível superior estão dispostos a pagar um valor maior que os produtores sem nível superior.

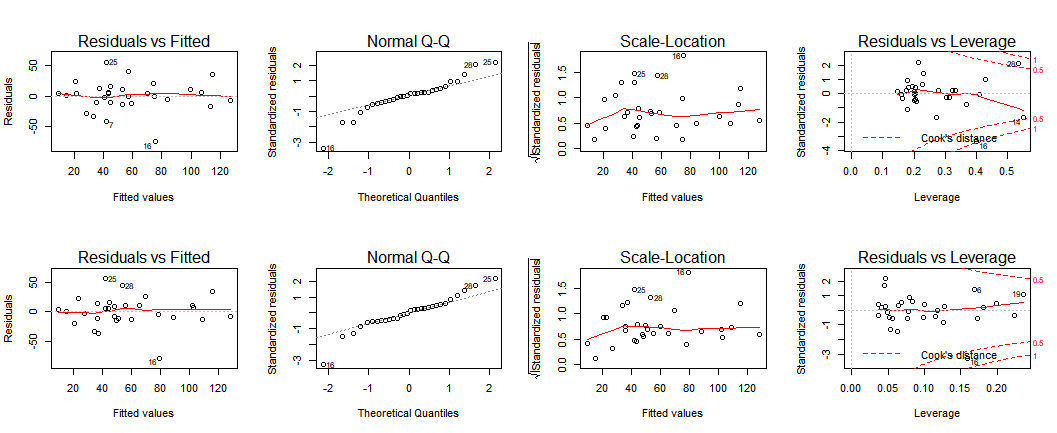
**Atividade 2: Analise estatística em R**

- Uma análise feita em R usando modelos lineares (com código da função usada e fórmula)

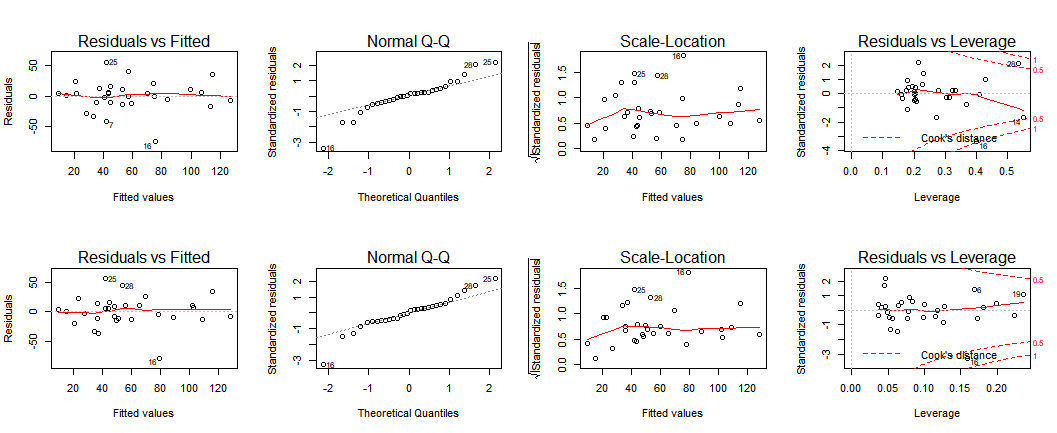
|  |
| --- |
| #Criar um modelo com todas as variaveis  modelo1 <- lm(data = valoracao\_ambiental,  formula = DAP ~ renda  + prod + sexo + grad + custo + idade)  #Observação do modelo1  summary(modelo1)  Call:  lm(formula = DAP ~ renda + prod + sexo + grad + custo + idade,  data = valoracao\_ambiental)  Residuals:  Min 1Q Median 3Q Max  -75.257 -9.922 2.296 11.461 56.354  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 12.8605 58.2487 0.221 0.8273  renda 6.8997 3.2668 2.112 0.0463 \*  prod 0.5176 0.2570 2.014 0.0564 .  sexomulher 9.6625 14.1114 0.685 0.5007  gradens.med -3.3664 13.1204 -0.257 0.7999  gradgraduado 1.5197 14.5872 0.104 0.9180  custo -0.1696 3.2331 -0.052 0.9586  idade -0.2535 0.5615 -0.451 0.6561  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 28.71 on 22 degrees of freedom  Multiple R-squared: 0.6081, Adjusted R-squared: 0.4834  F-statistic: 4.877 on 7 and 22 DF, p-value: 0.001924  #Funcao Step para selecionar as melhores variaveis  #Scale = Erro padrao residual  step(modelo1, direction = "both", scale = 28.71^2)  Start: AIC=8  DAP ~ renda + prod + sexo + grad + custo + idade  Df Sum of Sq RSS Cp  - grad 2 90.6 18222 4.1068  - custo 1 2.3 18134 5.9996  - idade 1 168.0 18299 6.2006  - sexo 1 386.4 18518 6.4656  <none> 18131 7.9968  - prod 1 3342.5 21474 10.0520  - renda 1 3676.4 21808 10.4571  Step: AIC=4.11  DAP ~ renda + prod + sexo + custo + idade  Df Sum of Sq RSS Cp  - custo 1 1.0 18223 2.1081  - idade 1 146.6 18368 2.2847  - sexo 1 408.6 18631 2.6026  <none> 18222 4.1068  - prod 1 4054.2 22276 7.0253  - renda 1 4134.0 22356 7.1221  + grad 2 90.6 18131 7.9968  Step: AIC=2.11  DAP ~ renda + prod + sexo + idade  Df Sum of Sq RSS Cp  - idade 1 148.0 18371 0.2876  - sexo 1 420.6 18644 0.6183  <none> 18223 2.1081  + custo 1 1.0 18222 4.1068  - prod 1 4061.8 22285 5.0358  - renda 1 4464.3 22687 5.5242  + grad 2 89.4 18134 5.9996  Step: AIC=0.29  DAP ~ renda + prod + sexo  Df Sum of Sq RSS Cp  - sexo 1 344.6 18716 -1.2943  <none> 18371 0.2876  + idade 1 148.0 18223 2.1081  + custo 1 2.4 18368 2.2847  - renda 1 4391.4 22762 3.6153  - prod 1 4414.4 22785 3.6431  + grad 2 71.1 18300 4.2013  Step: AIC=-1.29  DAP ~ renda + prod  Df Sum of Sq RSS Cp  <none> 18716 -1.2943  + sexo 1 344.6 18371 0.2876  + idade 1 72.0 18644 0.6183  + custo 1 14.3 18701 0.6884  - renda 1 4121.7 22837 1.7062  + grad 2 98.1 18617 2.5866  - prod 1 5493.4 24209 3.3703  Call:  lm(formula = DAP ~ renda + prod, data = valoracao\_ambiental)  Coefficients:  (Intercept) renda prod  1.318 6.725 0.596  #A Renda e a Produção foram selecionadas como melhores variaveis explicativas  #Gerar outro modelo com as duas variaveis  modelo2 <- lm(data = valoracao\_ambiental,  formula = DAP ~ renda + prod)  #Observar o novo modelo com as duas variaveis  summary(modelo2)  Call:  lm(formula = DAP ~ renda + prod, data = valoracao\_ambiental)  Residuals:  Min 1Q Median 3Q Max  -79.127 -12.001 1.903 10.597 55.905  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 1.3175 13.3518 0.099 0.92212  renda 6.7250 2.7578 2.438 0.02161 \*  prod 0.5960 0.2117 2.815 0.00899 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 26.33 on 27 degrees of freedom  Multiple R-squared: 0.5955, Adjusted R-squared: 0.5655  F-statistic: 19.87 on 2 and 27 DF, p-value: 4.942e-06  #Observar o diagnostico dos modelos  par(mfrow = c(2,4))  plot(modelo1)  plot(modelo2) |

- Uma visualização do modelo e de seu diagnóstico

Modelo 1, com todas variáveis:



Modelo 2, com as variáveis mais significativas para o modelo:



Legenda: R – Renda; P – Produção(kg); M – Produtor do sexo feminino; EM – Produtor com ensino médio; ES – Produtor com ensino superior; C – Custo de produção; Id – Idade.

- Uma explicação de suas conclusões e interpretações do modelo.

Os modelos testados apresentaram baixo grau de correlação entre o perfil dos produtores com a seus lances de disposição a pagar pela criação de um fundo de investimento para conservação da bacia hidrográfica do Rio Jaguaribe. Entre as variáveis que compunham o perfil dos produtores, apenas a renda e a produtividade dos produtores foram capazes de explicar a vontade dos mesmos de pagar. A baixa capacidade de explicação dos modelos está associada com a não normalidade dos resíduos ou a não adequação dos modelos lineares com os dados amostrados.