

# Modelo MANOVA

```
## Função que carrega e instala pacotes
ipak <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}
pkg <- c("tidyverse","plyr","plotrix","car") #plotrix para "plotic" car para "qqplot"
ipak(pkg)

## Loading required package: tidyverse

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():      dplyr, stats

## Loading required package: plyr

## -----

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## -----

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## The following object is masked from 'package:purrr':
##
##   compact

## Loading required package: plotrix

## Loading required package: car

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##   recode
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      some
```

```
## tidyverse      plyr      plotrix      car
```

```
##      TRUE      TRUE      TRUE      TRUE
```

```
##calcula coeficiente de correlacao(%)
```

```
cv <- function(x){
```

```
  100 * sd(x)/mean(x)
```

```
}
```

```
#lendo as funcoes que o caio disponibilizou para MANOVA (mudei uma coisa lá)
```

```
source("codigoscaioMANOVA.r")
```

```
#nomes para usar nas tabelas
```

```
inames <- c("Espécie","Comprimento da Asa","Largura da Asa","Comprimento 3° palpo","Largura 3° palpo",
```

```
#Identificação das colunas do Banco de dados
```

```
dadosmoscas <- read.table(file = "dadosmoscas.txt",col.names = c("Esp","Comp_Asa","Larg_Asa","Comp_3_pa
```

```
dadosmoscas$Esp=as.factor(dadosmoscas$Esp )#definindo o tipo da especie como fator
```

```
#glimpse(dadosmoscas) #Verificando os tipos de variaveis.
```

```
#Pegando a coluna das espécies (grupos) para fazer o vetor identificador de grupos
```

```
v.especies <- cbind(as.numeric(dadosmoscas[,1]))
```

```
#tirando a primeira coluna para fazer a matriz de dados
```

```
m.dados <- dadosmoscas[,-1]
```

```
#Criando um vetor com as quantidades de observações de cada grupo, constam 35 de cada grupo
```

```
v.n_especies <- rbind(35,35)
```

```
#Identificação da quantidade de grupos
```

```
k.G <- 2
```

```
#Identificação da quantidade de variáveis
```

```
k.P <- 7
```

```
#Teste de igualdade das matrizes de covariância entre os grupos
```

```
#resuldado da função
```

```
test.igual.matriz <-Box.teste.Igual.MCov(as.matrix(m.dados),v.especies,v.n_especies,k.G)
```

```
## Estatística do Teste: 64.53906
```

```
## nível descritivo: 0.0001038709
```

```
## Matrizes de Covariâncias por grupo:
```

```
##      grupo  Comp_Asa  Larg_Asa  Comp_3_palpo  Larg_3_palpo
```

```
## Comp_Asa      1 40.726050 11.7168067      2.3252101  2.199159664
```

```
## Larg_Asa      1 11.716807  7.4924370      1.8268908  1.839495798
```

```
## Comp_3_palpo  1  2.325210  1.8268908      4.8285714 -0.784873950
```

```
## Larg_3_palpo      1  2.199160  1.8394958  -0.7848739  3.374789916
## Comp_4_palpo      1  6.263025  3.2613445   0.6126050  1.696638655
## Comp_12_seg_ant   1  1.878151  0.3151261  -0.3655462  0.168067227
## Comp_13_seg_ant   1  1.663866  0.2983193   0.1680672  0.004201681
## Comp_Asa          2 31.290756 17.4731092   9.8302521  5.150420168
## Larg_Asa          2 17.473109 25.7848739   3.7008403  4.144537815
## Comp_3_palpo      2  9.830252  3.7008403   8.0453782  2.169747899
## Larg_3_palpo      2  5.150420  4.1445378   2.1697479  2.702521008
## Comp_4_palpo      2 12.882353  8.9411765   2.6176471  3.117647059
## Comp_12_seg_ant   2  2.973950  1.7915966   0.7873950  0.378991597
## Comp_13_seg_ant   2  3.633613  1.5689076   1.1739496  0.454621849
##
##      Comp_4_palpo Comp_12_seg_ant Comp_13_seg_ant
## Comp_Asa      6.26302521      1.87815126      1.663865546
## Larg_Asa      3.26134454      0.31512605      0.298319328
## Comp_3_palpo   0.61260504     -0.36554622      0.168067227
## Larg_3_palpo   1.69663866      0.16806723      0.004201681
## Comp_4_palpo   6.24033613     -0.01680672     -0.050420168
## Comp_12_seg_ant -0.01680672      0.84033613      0.638655462
## Comp_13_seg_ant -0.05042017      0.63865546      0.798319328
## Comp_Asa      12.88235294      2.97394958      3.633613445
## Larg_Asa       8.94117647      1.79159664      1.568907563
## Comp_3_palpo    2.61764706      0.78739496      1.173949580
## Larg_3_palpo    3.11764706      0.37899160      0.454621849
## Comp_4_palpo    21.29411765      1.14705882      1.323529412
## Comp_12_seg_ant  1.14705882      1.58487395      1.189915966
## Comp_13_seg_ant  1.32352941      1.18991597      1.181512605
```

```
#matriz sigma P (não lembro o que é exatamente)
```

```
m.sigma.P <- test.igual.matriz$Sigma.P
```

```
#estatística do teste
```

```
est.test.igual.m.cov <- test.igual.matriz$est.test
```

```
#p-valor do teste
```

```
p.valor.igual.m.cov <- test.igual.matriz$p.valor
```

```
#Teste e resultados MANOVA
```

```
fit.manova <- manova(as.matrix(m.dados) ~ v.especies)
```

```
#Aqui é apresentado testes anova para cada variável
```

```
anovas <- summary.aov(fit.manova)
```

```
#Aqui é o resultado das estatísticas de teste que ele passou para MANOVA
```

```
#OBS acho que vai ter que copiar os p-valores "na mão" pq não está fácil extrai-los
```

```
Wilks <- summary.manova(fit.manova,test="Wilks")
```

```
Pillai <- summary.manova(fit.manova,test="Pillai")
```

```
Hotelling <- summary.manova(fit.manova,test="Hotelling-Lawley")
```

```
Roy <- summary.manova(fit.manova,test="Roy")
```

```
# Parâmetros preditos pelo modelo
```

```
m.B <- matrix(coef(fit.manova),2,7) #2 Grupos e 7 variáveis
```

```
v.beta <- matrix(t(m.B))
```

```
m.X <- model.matrix(fit.manova)
```

```
m.Sigmabeta <- kronecker(solve(t(m.X)%*%m.X),m.sigma.P)
```

```
#Médias preditas pelo modelo com os respectivos Erro padrão e IC's
#Lembre que algumas matrizes usadas aqui foram definidas no código anterior à isto (logo acima)
```

```
#k.posicao_Var é o índice "k" do \mu_k que você quer prever
MedidasPreditasModelo <- function(k.posicao_var){

m.x_mu <- matrix(data = rep(0,28), nrow = 2, ncol = 14)
m.x_mu[,k.posicao_var] <- 1
m.x_mu[2,k.posicao_var+7] <- 1

m.mu_pred <- m.x_mu%*%v.beta
m.ep_mu_pred <- sqrt(diag(m.x_mu%*%m.Sigmabeta%*%t(m.x_mu)))
m.IC_mu_pred <- cbind(m.mu_pred-1.96*m.ep_mu_pred,m.mu_pred+1.96*m.ep_mu_pred)

return(list(mu = m.mu_pred,ep = m.ep_mu_pred, IC = m.IC_mu_pred))
}
```

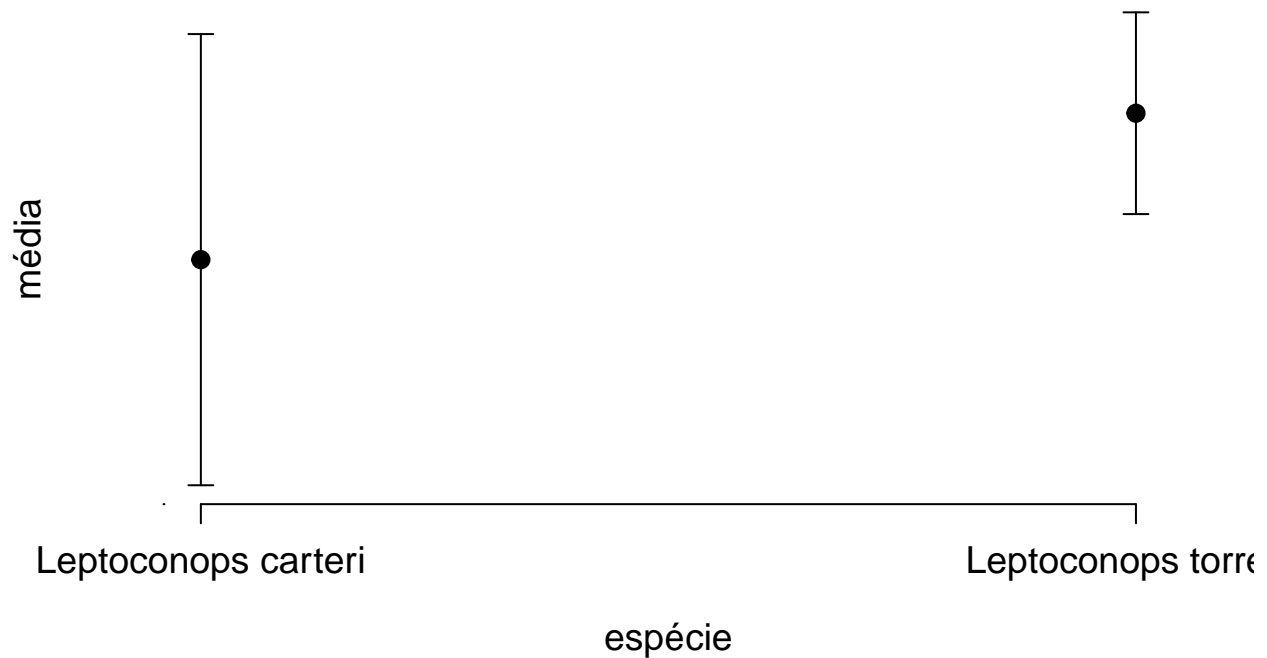
```
#Não coloquei o gráfico na função "MedidasPreditasModelo" pois a função plotCI não é do ggplot, ai não
```

```
#lembrar de mudar o título dos gráficos.
```

```
a <- MedidasPreditasModelo(k.posicao_var = 1)

plotrix::plotCI(a$mu,ui=a$IC[,2],li=a$IC[,1],axes=FALSE,xlab="espécie",ylab="média",pch=19,cex=1.2,cex.
axis(2,seq(4,7,0.2),cex.axis=1.2)
axis(1,1:2,c("Leptoconops carteri","Leptoconops torrens"),cex.axis=1.2)
```

## Comprimento da Asa



*#Aqui os testes CBU=M vou fazer de uma maneira geral pois depende da análise descritiva . AQUI COMPARA-*

*#Para testar  $\alpha_{2i} = \alpha_{2j} = \alpha_{2k} = \alpha_{2s} = 0$   
#OBS caso for comparações de menor ordem é só ignorar os parâmetros restantes*

```
GeraMatriz_C_B_U <- function(k.ncomp,k.i,k.j,k.k,k.s) {
```

```
  #k.ncomp <- 3 #número de parâmetros \alphas a serem comparados à 0.
```

```
  #k.i <- 0 #mude aqui quando definir as comparações, isto define a posição
```

```
  #k.j <- 0 #mude aqui quando definir as comparações, isto define a posição
```

```
  #k.k <- 0 #mude aqui quando definir as comparações, isto define a posição
```

```
  #k.s <- 0 #mude aqui quando definir as comparações, isto define a posição
```

```
  m.C1 <- cbind(0,1) # para testar os \alpha's
```

```
  #Comparação de k.ncomp \alpha's com zero e são 7 variáveis
```

```
  m.U1 <- matrix(data = c(rep(0,k.ncomp*7)), nrow = 7, ncol = k.ncomp)
```

```
  m.U1[k.i,1] <- 1
```

```
  if(k.ncomp >= 2) m.U1[k.j,2] <- 1
```

```
  if(k.ncomp >= 3) m.U1[k.k,3] <- 1
```

```
  if(k.ncomp >= 4) m.U1[k.s,4] <- 1
```

```
  m.M <- rbind(rep(0,k.ncomp))
```

```

  return(list(C = m.C1, U = m.U1, M = m.M))
}

```

*#Um exemplo abaixo de como utilizar a função, aqui está comparando simultaneamente se as espécies de mo.*

```

m.teste1 <- GeraMatriz_C_B_U(k.ncomp = 2,1,2)
resultados_teste1 <- TesteF.CBU.M(fit.manova,m.sigma.P,k.P,k.G,m.teste1$C,m.teste1$U,m.teste1$M)

```

```

## Estatística Qui-quadrado = 4.24
## pvalor = 0.1202
## Matriz C :
##      [,1] [,2]
## [1,]    0    1
## Matriz U :
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
## [3,]    0    0
## [4,]    0    0
## [5,]    0    0
## [6,]    0    0
## [7,]    0    0
## Matriz M :
##      [,1] [,2]
## [1,]    0    0

```

```
resultados_teste1$estatqui
```

```

##      [,1]
## [1,] 4.236919

```

```

legendaaux <- rep(" ",length(inames)-1)
for (i in 2:length(inames)) {
  aux <- "Gráficos para os resíduos referentes à variável"
  aux2 <- as.character(inames[i])
  legendaaux[i-1] <- paste(aux,aux2,sep = " ")
}
legenda <- legendaaux[1]

```

*#Esta função é só pra não ficar copiando, da pra colocar fora do código e chamar com o source.  
#Por causa da legenda, não vai adiantar essa função por causa da legenda*

*#FUNÇÃO DESATIVADA*

```
Seteplots_Residuos <- function(mY,mresult,var,typeresid,wplot){
```

```

  # 1 Variável
  var<-1; gen.graf.resid(mY,mresult,var,typeresid,wplot)

  var<-2; gen.graf.resid(mY,mresult,var,typeresid,wplot)
  # 3 Variável
  var<-3; gen.graf.resid(mY,mresult,var,typeresid,wplot)
  # 4 Variável
  var<-4; gen.graf.resid(mY,mresult,var,typeresid,wplot)
  # 5 Variável

```

```

var<-5; gen.graf.resid(mY,mresult,var,typeresid,wplot)
# 6 Variável
var<-6; gen.graf.resid(mY,mresult,var,typeresid,wplot)
# 7 Variável
var<-7; gen.graf.resid(mY,mresult,var,typeresid,wplot)
}

#Os gráficos são bem parecidos com os multivariados. Mas aqui usa "Resíduo studentizado (RS)"
mY <- as.matrix(m.dados)
mresult<- fit.manova
typeresid <- "univariate"
wplot <- "diagnostics"
#par(mfrow =c(4,2))
Seteplots_Residuos(mY,mresult,var,typeresid,wplot)

#DESATIVADA
# Resíduos univariados
#Os gráficos são bem parecidos com os multivariados. Mas aqui usa "Resíduo studentizado (RS)"

diag_residuos <- function(){
  var<-1; gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")
  var <- var + 1
  legenda <- legendaaux[var]
}

var = 0
var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

var<-1+var
legenda <- legendaaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")

```

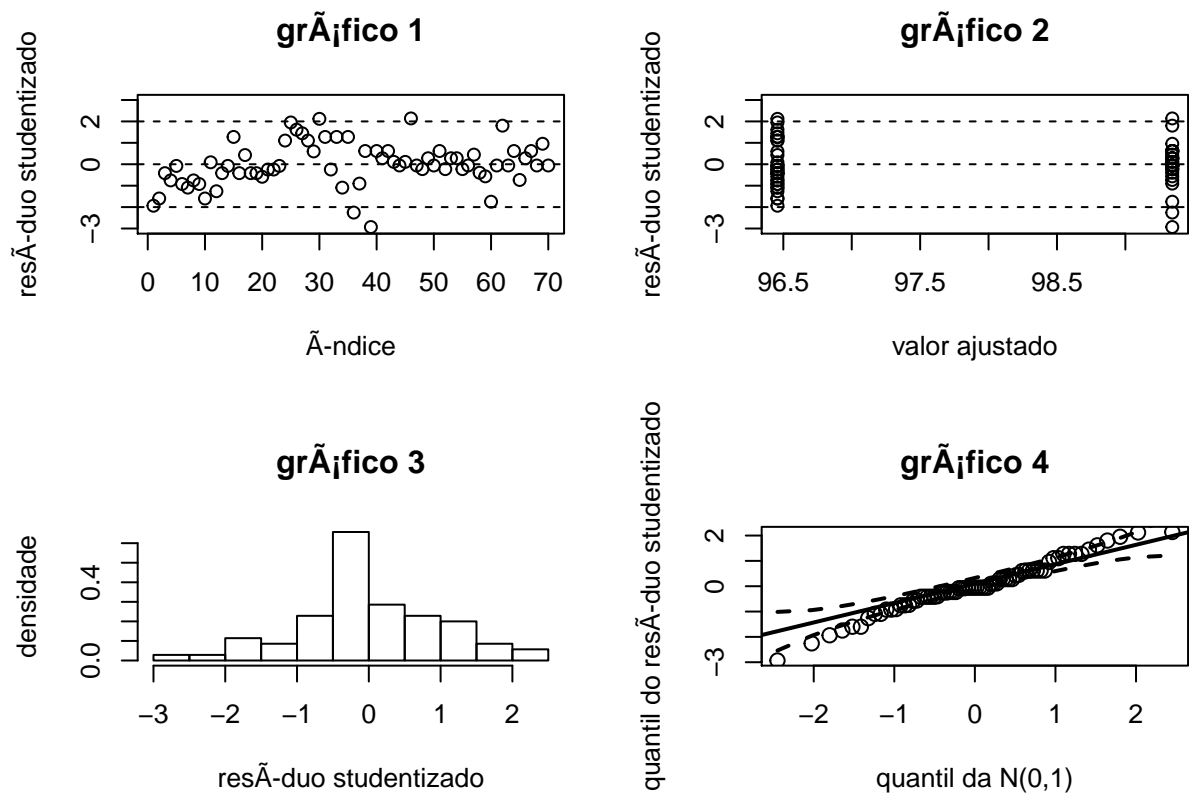


Figure 1: Gráficos para os resíduos referentes à variável Comprimento da Asa



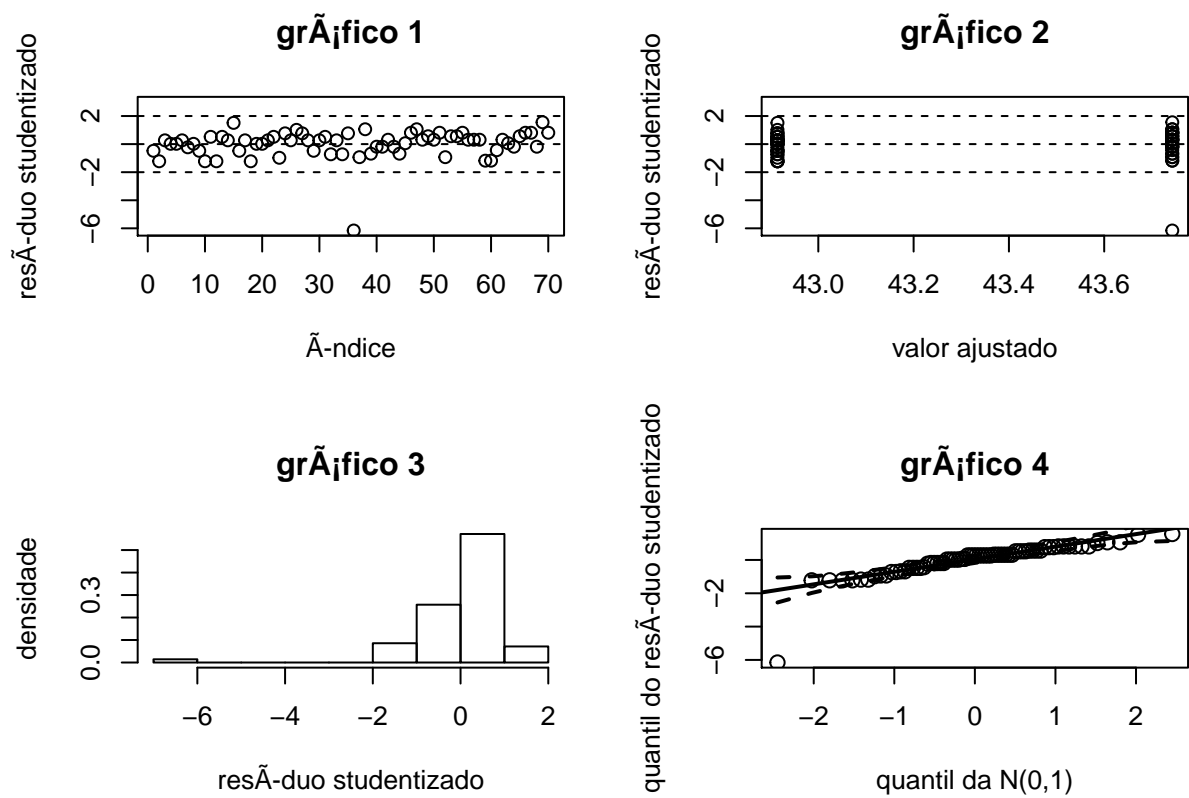


Figure 2: Gráficos para os resíduos referentes à variável Largura da Asa

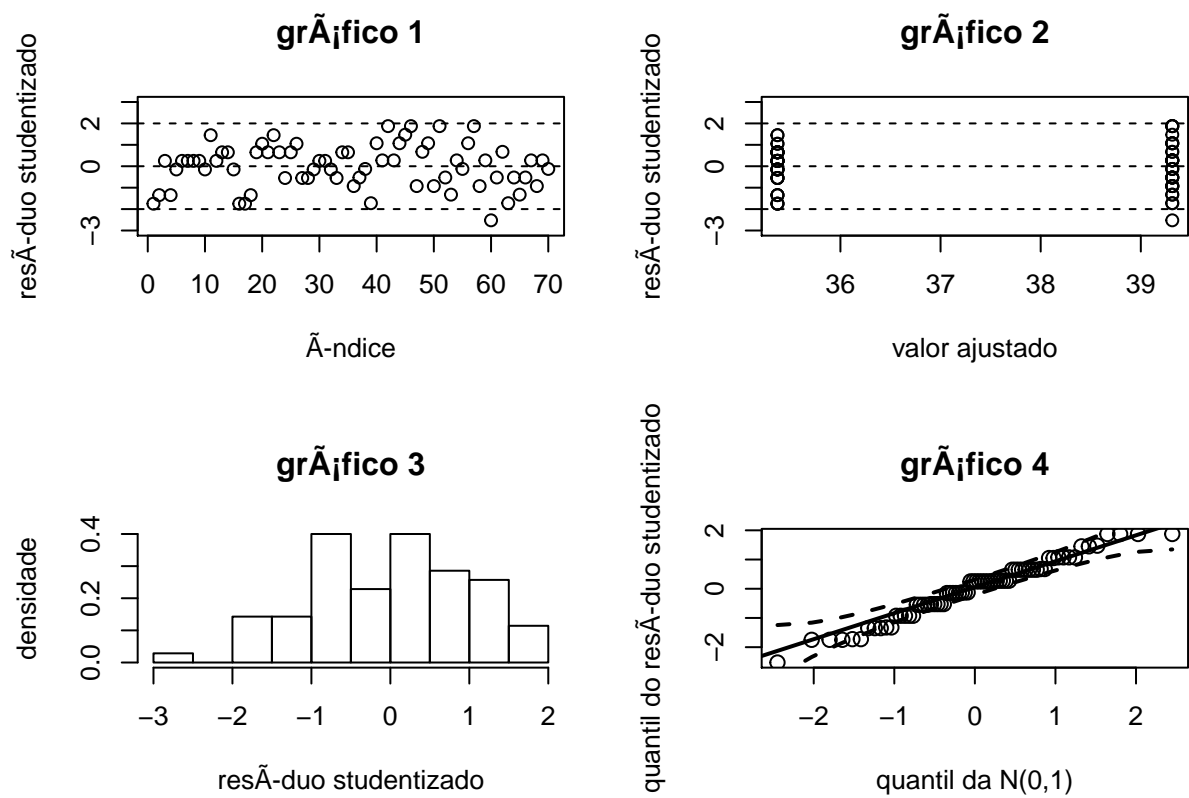


Figure 3: Gráficos para os resíduos referentes à variável Comprimento 3º palpo

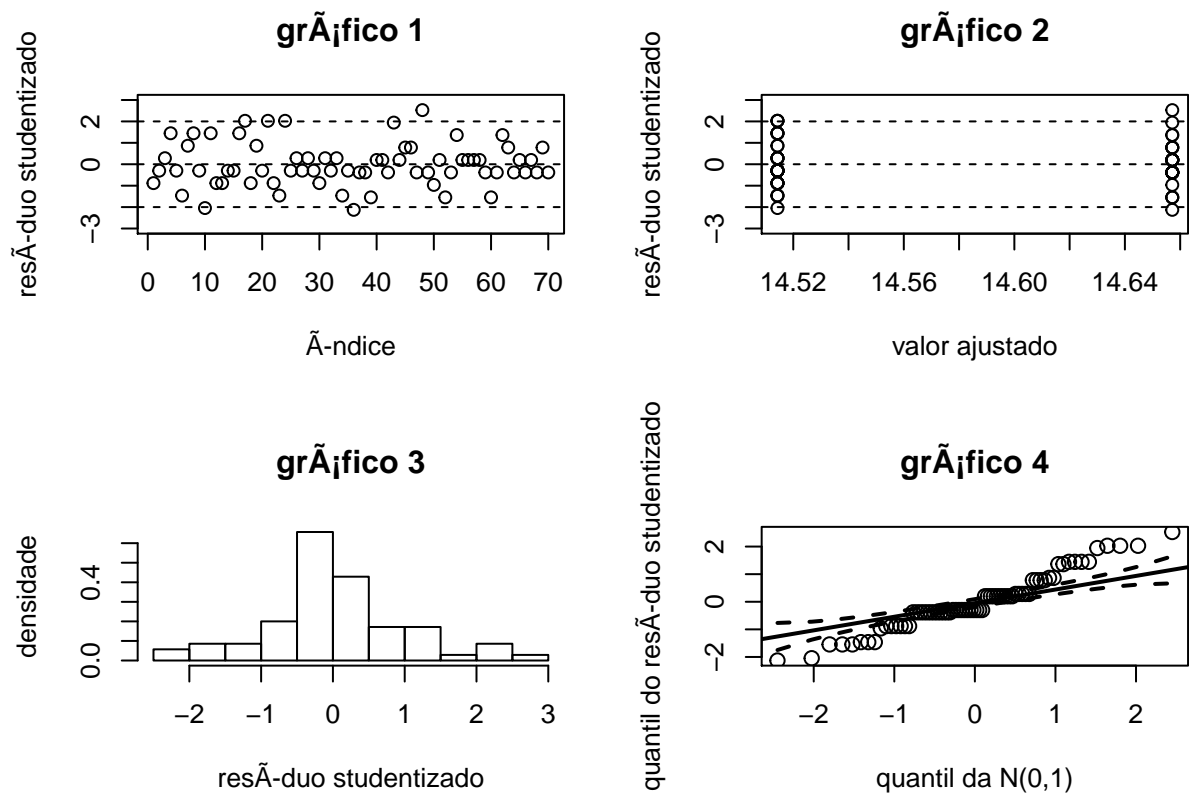


Figure 4: Gráficos para os resíduos referentes à variável Largura 3° palpo

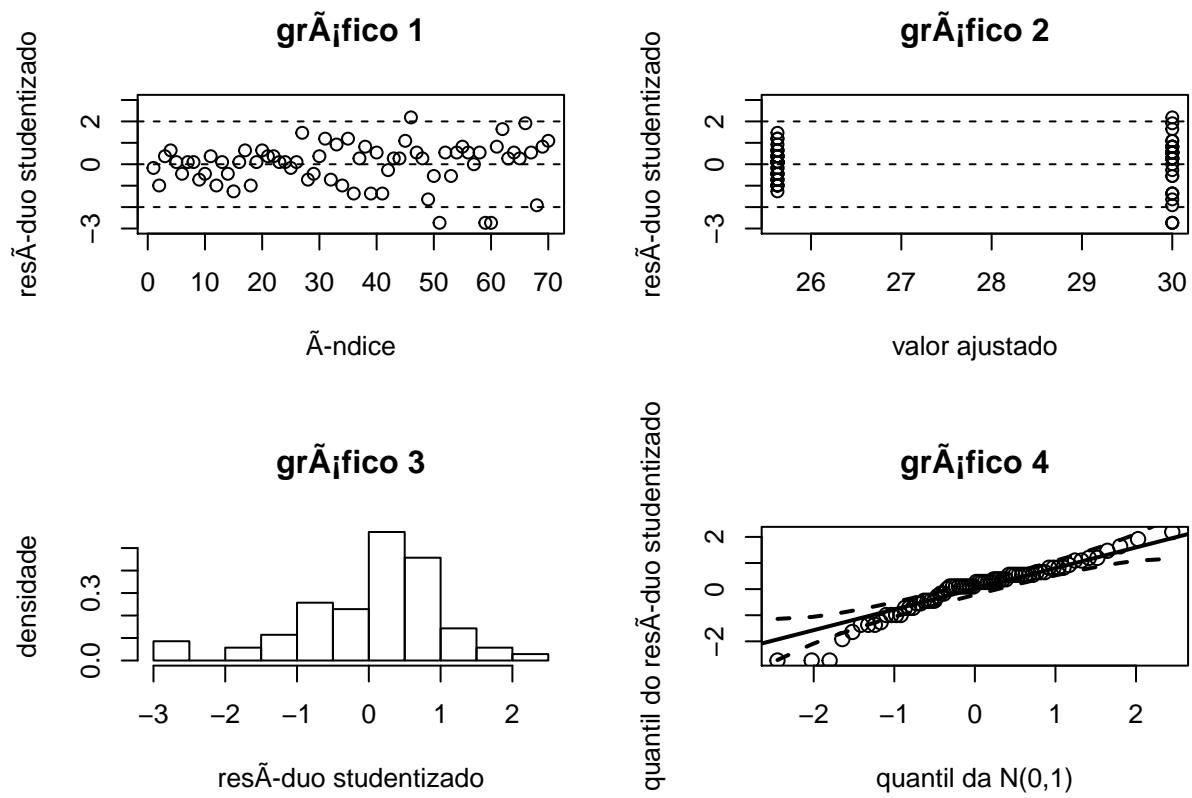


Figure 5: Gráficos para os resíduos referentes à variável Comprimento 4º palpo

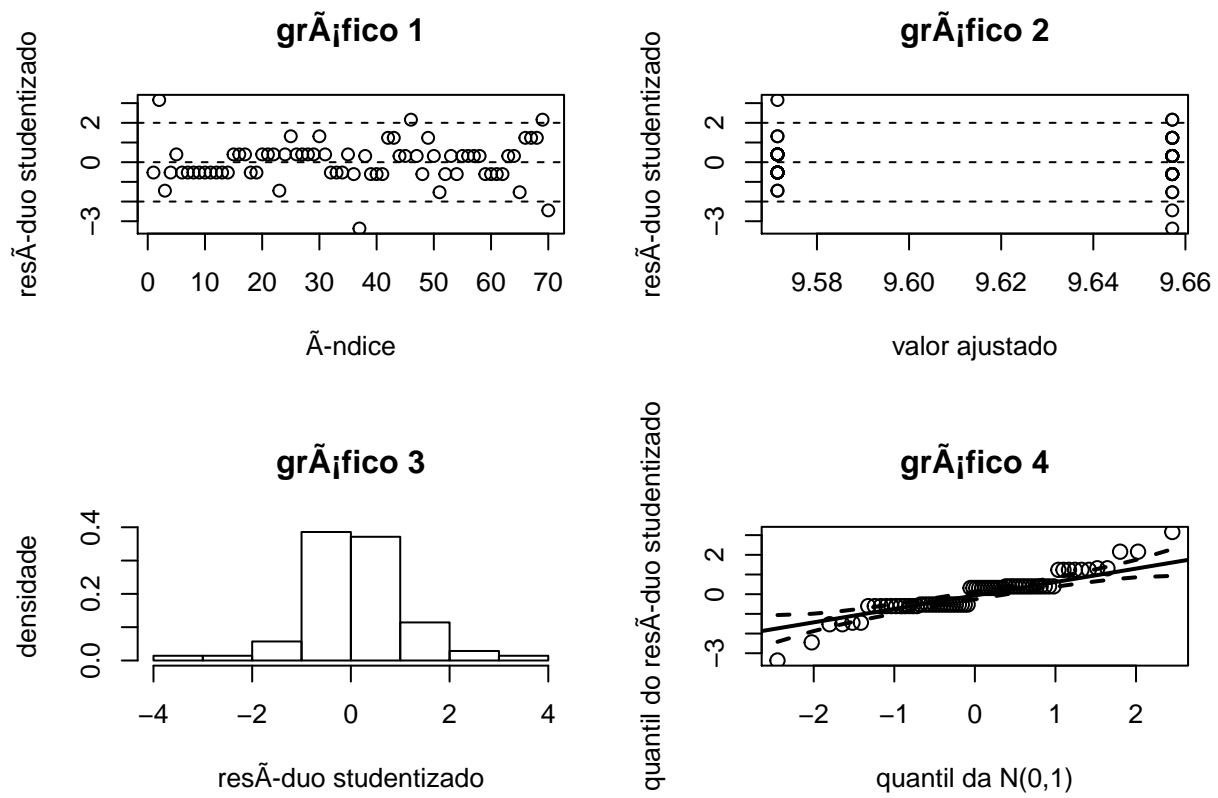


Figure 6: Gráficos para os resíduos referentes à variável Comprimento 12º segmento da antena

```
var<-1+var
legenda <- legendaux[var]

gen.graf.resid(as.matrix(m.dados),fit.manova,var,"univariate","diagnostics")
```

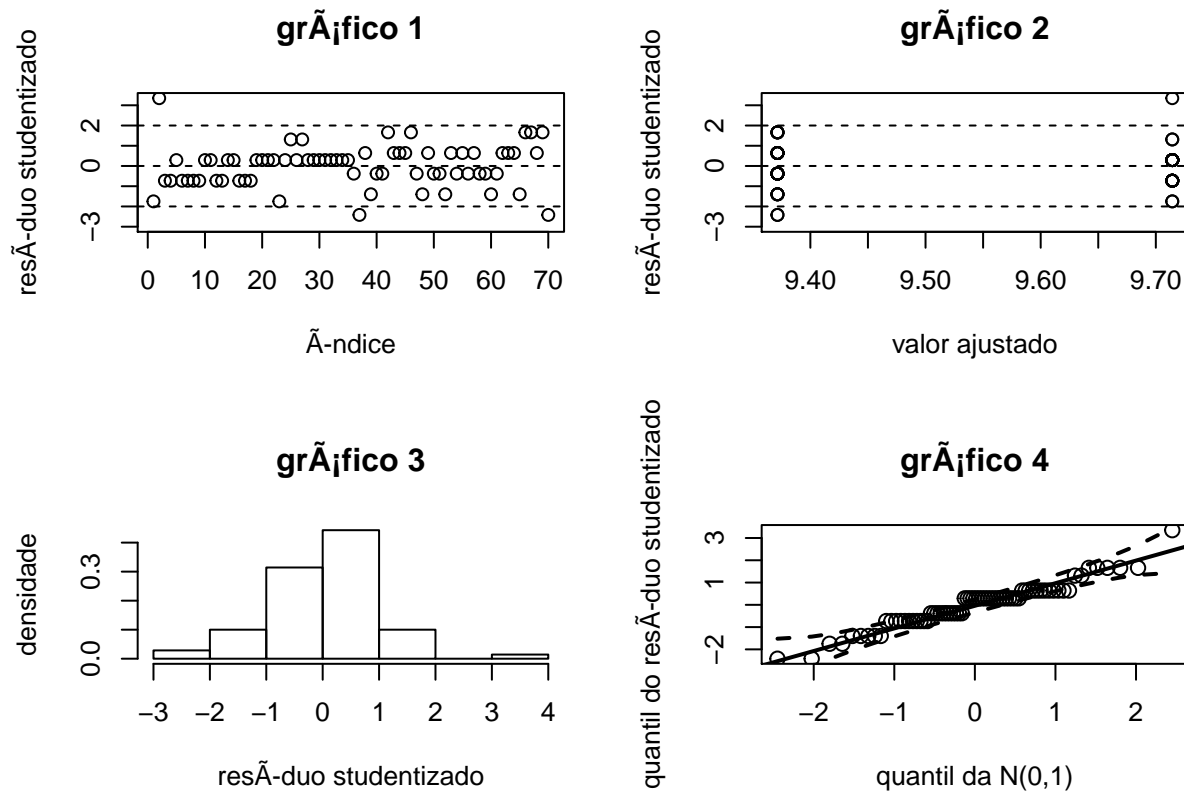


Figure 7: Gráficos para os resíduos referentes à variável Compimento 13° segmento da antena

```
#Acho que não vai precisar desses

# Zoom nos envelopes
wplot <- "envelope"

par(mfrow =c(4,2))
Seteplots_Residuos(mY,mresult,var,typeresid,wplot)

# Resíduos multivariados
#Os gráficos são bem parecidos com os univariados. Mas aqui usa "Resíduo studentizado (multivariado) (R

typeresid <- "multivariate"
wplot <- "diagnostics"

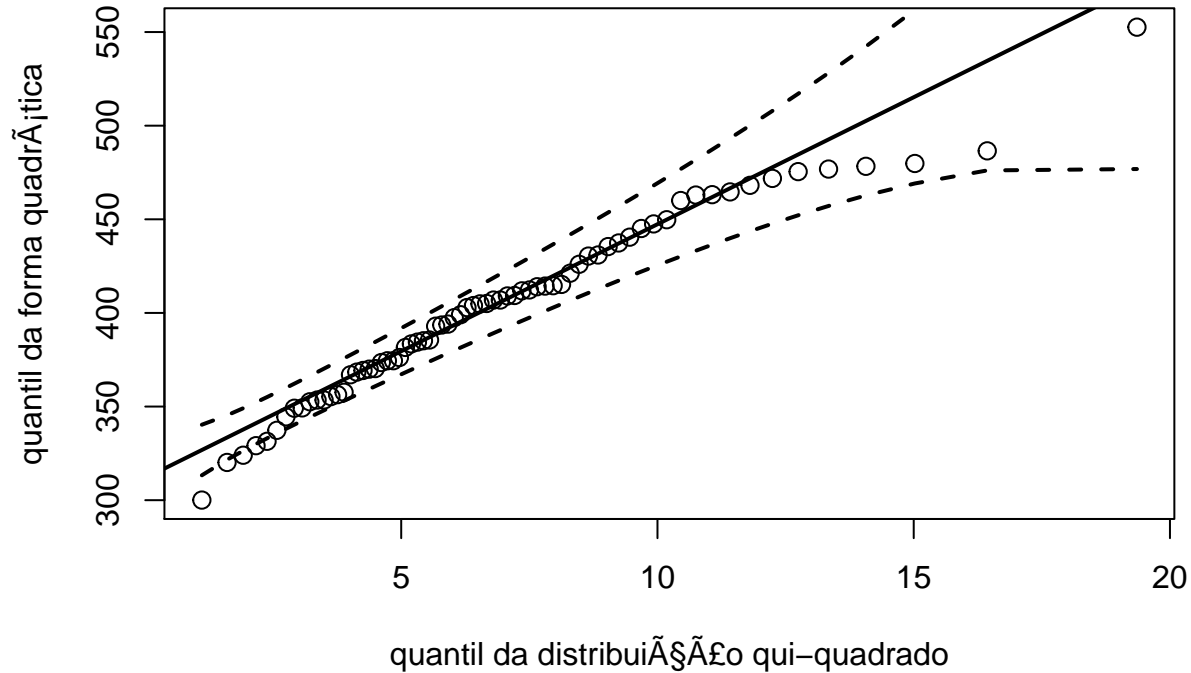
Seteplots_Residuos(mY,mresult,var,typeresid,wplot)

# Zoom nos envelopes
wplot <- "envelope"
```

```
Seteplots_Residuos(mY,mresult,var,typeresid,wplot)

# Resíduo baseado na distância Mahalanobis

gen.graf.resid.quad.form(as.matrix(m.dados),fit.manova)
```



$$Y_{ijk} = \mu_k + \alpha_{ik} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N_k(0, \Sigma),$$

$i = 1, 2$  (espécie, 1 - *Leptoconops torrens*, 2 - *Leptoconops carteri*),  $j = 1, 2, \dots, 35$  (moscas) e  $k = 1, \dots, 7$  (variável, 1 - Comprimento da Asa, 2 - Largura da Asa, 3 - Comprimento 3º palpo, 4 - Largura 3º palpo, 5 - Comprimento 4º palpo, 6 - Comprimento 12º segmento da antena, 7 - Comprimento 13º segmento da antena), em que  $\alpha_{1k} = 0, k = 1, \dots, 7$ .

### 3.1 Análise dos resíduos

$X <-$  é o número do primeiro gráfico de resíduos

A fim de avaliar a validade das suposições de normalidade multivariada dos dados considerando as espécies (consequentemente normalidade univariada) e homocedasticidade multivariada entre as espécies (consequentemente homocedasticidade univariada), podemos observar as figuras (X) a (X+7) que apresentam gráficos para os resíduos studentizados para cada uma das 7 variáveis, assim como a figura (X+8) que apresenta o gráfico de envelopes baseado na distância de Mahalanobis (veja Azevedo, 2017). A partir da observação destes gráficos, pode-se identificar muitos comportamentos e tendências não esperadas, as quais podemos destacar o comportamento apresentado no gráfico 4 das figuras (X), (X+3), (X+5) e (X+6) tendo muitos pontos fora dos limites das bandas de confiança, nas figuras (X+1), (X+2) parece existir uma pequena tendência nos valores dos resíduos e na figura (X+4) apresenta muitos pontos com quantis baixos fora das

bandas de confiança. Adicionalmente, identificamos comportamento assimétrico negativo no gráfico 3 das figuras (X), (X+1), (X+2) e (X+4) e assimétrico positivo apresentando na figura (X+6). Dadas as observações referentes aos gráficos 1 e 4 das figuras (X) a (X+6) temos um forte indício de que a suposição de normalidade não é razoável para nenhuma das variáveis presentes no banco de dados. Observando o gráfico 2 das figuras (X) a (X+7), identificamos evidências de presença de heterocedasticidade dos dados nas figuras (X), (X+2), (X+3) de maneira mais leve e nas figuras (X+4), (X+5) e (X+6) de maneira mais acentuada, já para a figura (X+1) não nota-se, a menos de um valor extremo, a presença de indícios de heterocedasticidade. Não identificamos nenhum comportamento a ser destacado referente ao gráfico 1 das figuras (X) a (X+7). Na figura (X+8) observamos alguns valores fora das bandas de confiança para valores menores de quantis da forma quadrática, além disso, valores maiores de quantis da forma quadrática tendem a se apresentar abaixo da linha de referência baseada no quantil da distribuição qui-quadrado, deste modo temos indicações de que a suposição de normalidade multivariada dos dados não parece ser uma suposição razoável neste caso. Contudo, dadas as observações destacadas, temos que a única variável do banco de dados a qual não seria irrisuável supor normalidade e homocedasticidade dos dados seria a variável “Largura da Asa”, e todas as restantes apresentam ao menos um indício evidente da fuga destas suposições portanto não seria razoável supor normalidade e homocedasticidade multivariada neste caso, o que também fica evidente na figura (X+8), sendo assim, o modelo de análise de variância multivariada não apresentou um ajuste adequado aos dados aqui analisados e se é necessário procurar técnicas alternativas para realizar uma análise adequada ao banco de dados. Dado o nosso contexto acadêmico, iremos continuar com as análises dos resultados para elaborar a conclusão do presente trabalho.