

# Códigos utilizados no laboratório 8

Camille Menezes e Michel Miler

Junho de 2023

## 1 Atividade 1

```
library(GGally)
library(tidyverse)
library(patchwork)
library(lmtest)
library(MASS)
library(lindia)
library(dplyr)
library(gtsummary)
library(rigr)
library(olsrr)
library(corrplot)
library(broom)
library(faraway)

dados <- read.table("Lab08q01.txt", header = T)

dados <- dados|>
  mutate(RENDAC = RENDA-mean(RENDA))

my_fn <- function(data, mapping, ...){
  p <- ggplot(data = data, mapping = mapping) +
    geom_point() +
    geom_smooth(method=lm,
                fill="blue", color="blue", ...)
  p
}

ggpairs(dados,
        upper = list(continuous = "cor"),
        lower = list(continuous = my_fn),
        axisLabels="none")

summary(dados)

boxp <- function(y,name){
  ggplot(dados) +
    aes(x=1,y=dados[,y]) +
    geom_boxplot() +
    stat_summary(fun=mean, geom="point", shape=8, size=3) +
    scale_x_discrete("",breaks = NULL) +
```

```

    geom_text(aes(label=ifelse( dados[,y] > as.numeric(quantile(dados[,y])[4] + 1.5*(quantile(dados[,y],
        as.character(rownames(dados))),')),
        hjust=-0.3,vjust=0) +
    labs(y=name) +
    theme_minimal() +
    if (y == 2){
        annotate("text", x = Inf, y = Inf, hjust = 1, vjust = 1,
            colour = "black", label = paste(" = Média"))
    }
    else{
        theme_minimal()
    }
}

```

```

boxp(1, "Renda") | boxp(2, "Tempo")

```

```

#
models <- lm(TEMPO~RENDA, data=dados)
summary(models)

models2 <- lm(TEMPO~RENDA+I(RENDA^2), data=dados)
summary(models2)

models3 <- lm(TEMPO~RENDA+I(RENDA^2)+I(RENDA^3), data=dados)
summary(models3)

```

```

ggplot(dados) +
  aes(x=RENDA, y=TEMPO)+
  geom_point() +
  geom_line(aes(y=models$fitted.values,color="Linear"))+
  geom_line(aes(y=models2$fitted.values,color="Quadrático"))+
  geom_line(aes(y=models3$fitted.values,color="Cúbico"))+
  scale_color_manual(name="",values=c("Linear" = "red","Quadrático" = "green","Cúbico" = "blue"))+
  theme_minimal()

```

```

#
model <- lm(TEMPO~RENDAC, data=dados)
summary(model)

model2 <- lm(TEMPO~RENDAC+I(RENDAC^2), data=dados)
summary(model2)

model3 <- lm(TEMPO~RENDAC+I(RENDAC^2)+I(RENDAC^3), data=dados)
summary(model3)

```

```

ggplot(dados) +
  aes(x=RENDAC, y=TEMPO)+
  geom_point() +
  geom_line(aes(y=models$fitted.values,color="Linear"))+

```

```

geom_line(aes(y=models2$fitted.values,color="Quadrático"))+
geom_line(aes(y=models3$fitted.values,color="Cúbico"))+
scale_color_manual(values=c("Linear" = "red","Quadrático" = "green","Cúbico" = "blue"))+
theme_minimal()

#colinearidade
X <- model.matrix(models3)
tabela <- vif(X)
tabela <- c(1-1/vif(X),tabela)

tabela <- matrix(tabela,nrow=4,ncol=2)
tabela <- data.frame(tabela)
colnames(tabela) <- c('Rj2','VIF')

X <- model.matrix(model3)
tabela <- vif(X)
tabela<-c(1-1/vif(X),tabela)

tabela = matrix(tabela,nrow=4,ncol=2)
tabela <- data.frame(tabela)
colnames(tabela) <- c('Rj2','VIF')
round(tabela,2)
tabela

#TRV
tab <- augment(model3)
lrtest(model3, model2)

k <- tidy(anova(model3))
l <- tidy(anova(model2))
m <- tidy(anova(model))

SqRegc <- sum(k$sumsq[-4])
SqResc <- k$sumsq[4]

SqRegr = m$sumsq[1]

Fr = (SqRegc - SqRegr)*(nrow(dados)-4)/(SqResc*2)

1-pf(Fr,2,nrow(dados)-4)

SqRegr = sum(l$sumsq[-3])

Fr = (SqRegc - SqRegr)*(nrow(dados)-4)/(SqResc*1)

1-pf(Fr,1,nrow(dados)-4)

```

```

# Análise dos resíduos
shapiro.test(model2$residuals)
qqtest(model2)
dwtest(model2)

a <- ggplot(model2)+
  geom_point(aes(x=.fitted,y=studres(model2)))+
  labs(x="Valores preditos \n A",y="Resíduos estudentizados")+
  theme_minimal()

b <- ggplot(model2, aes(sample=studres(model2)))+
  geom_qq()+
  geom_qq_line(color="red")+
  labs(x="Quantil teórico \n B", y="Resíduos estudentizados")+
  theme_minimal()

c <- ggplot(model2)+
  geom_histogram(aes(x=.resid),binwidth=2,color="black", fill="white")+
  labs(y="Frequência",x="Resíduos \n C")+
  theme_minimal()

a | b | c

#18,5,19
ols_plot_cooksd_bar(model2)

ols_plot_dfbetas(model2)

ols_plot_dffits(model2)

ols_plot_resid_stud(model2)

ols_plot_resid_stand(model2)

ols_plot_resid_lev(model2)

ols_plot_resid_stud_fit(model2)

ols_plot_resid_pot(model2)

```

## 2 Atividade 2

```
df <- prostate
names(df)
df <- df[, -c(5)]
```

```
df <- df |>
  rename(lvolca = "lcavol",
         lpeso = "lweight",
         idade = "age",
         lhpv = "lbph",
         lpc = "lcp",
         pg45 = "pgg45",
         laep = "lpsa")
```

```
my_fn <- function(data, mapping, ...){
  p <- ggplot(data = data, mapping = mapping) +
    geom_point() +
    geom_smooth(method=lm,
                fill="blue", color="blue", ...)
  p
}
```

```
ggpairs(df,
        upper = list(continuous = "cor"),
        lower = list(continuous = my_fn),
        axisLabels="none")
```

```
boxp <- function(y){
  ggplot(df) +
    aes(x=1,y=df[,y]) +
    geom_boxplot() +
    stat_summary(fun=mean, geom="point", shape=8, size=3) +
    scale_x_discrete("", breaks = NULL) +
    #geom_text(aes(label=ifelse( df[,y] > as.numeric(quantile(df[,y])[4] + 1.5*(quantile(df[,y])[4] -
    #                                as.character(rownames(df)),'')),
    #                                hjust=-0.3,vjust=0) +
    labs(y=names(df[y])) +
    theme_minimal() +
    if (y == 7){
      annotate("text", x = Inf, y = Inf, hjust = 1, vjust = 1,
               colour = "black", label = paste(" Média"))
    }
    else{
      theme_minimal()
    }
}
```

```
boxp(1) / boxp(2) / boxp(3) |
boxp(4) / boxp(5) / boxp(6) |
boxp(7) / boxp(8) / boxp(9)
```

```

# modelo completo
fit <- lm(laep ~ ., data = df)
summary(fit)

MyAnova <- function(modelo, nd){
  m1 <- modelo
  np <- dim(anova(m1))[1]
  SQReg <- round(sum(anova(m1)$"Sum Sq"[1:(np-1)]), nd)
  glReg <- sum(anova(m1)$"Df"[1:(np-1)])
  SQRes <- round(anova(m1)$"Sum Sq"[np], nd)
  glRes <- anova(m1)$"Df"[np]
  SQTotal <- round(SQReg + SQRes, nd)
  glTotal <- glReg + glRes
  QMReg <- round(SQReg/glReg, nd)
  QMRes <- round(SQRes/glRes, nd)
  MyF <- round(QMReg/QMRes, nd)
  vpF <- ifelse(pf(MyF, glReg, glRes, lower.tail = F) < 0.0001, "<0.001",
               roud(pf(MyF, glReg, glRes, lower.tail = F), nd))
  ncolumas <- c("Fonte de Variação", "SQ", "gl", "F", "valor p")
  Tanova <- data.frame(FV = c("Regressão",
                             "Resíduos",
                             "Total"),
                      gl = c(glReg, glRes, glTotal),
                      SQ = c(SQReg, SQRes, SQTotal),
                      QM = c(QMReg, QMRes, " "),
                      Est.F = c(MyF, " ", " "),
                      valor.p = c(vpF, " ", " ")
  )
  Tanova
}

MyAnova(fit,2)
anova(fit)

# Seleção de modelos -----
# a
anova(fit,lm(laep ~ lvolca+lpeso+idade+lhpb+lpc,data=df))
anova(fit,lm(laep ~ lvolca+lpeso+idade+lhpb+pg45,data=df))
anova(fit,lm(laep ~ lvolca+lpeso+idade+lpc+pg45,data=df))
anova(fit,lm(laep ~ lvolca+lpeso+lhpb+lpc+pg45,data=df))
anova(fit,lm(laep ~ lvolca+idade+lhpb+lpc+pg45,data=df))
anova(fit,lm(laep ~ lpeso+idade+lhpb+lpc+pg45,data=df))

#remove lpc
fit1 <- lm(laep ~ lvolca+lpeso+idade+lhpb+pg45,data=df)

anova(fit1, lm(laep ~ lvolca+lpeso+idade+lhpb,data=df))
anova(fit1, lm(laep ~ lvolca+lpeso+idade+pg45,data=df))
anova(fit1, lm(laep ~ lvolca+lpeso+lhpb+pg45,data=df))
anova(fit1, lm(laep ~ lvolca+idade+lhpb+pg45,data=df))
anova(fit1, lm(laep ~ lpeso+idade+lhpb+pg45,data=df))

```

```

# remove lhbp
fit2 <- lm(laep ~ lvolca+lpeso+idade+pg45,data=df)

anova(fit2,lm(laep ~ lvolca+lpeso+idade,data=df))
anova(fit2,lm(laep ~ lvolca+lpeso+pg45,data=df))
anova(fit2,lm(laep ~ lvolca+idade+pg45,data=df))
anova(fit2,lm(laep ~ lpeso+idade+pg45,data=df))

#remove idade
fit3 <- lm(laep ~ lvolca+lpeso+pg45,data=df)

anova(fit3,lm(laep ~ lvolca+lpeso,data=df))
anova(fit3,lm(laep ~ lvolca+pg45,data=df))
anova(fit3,lm(laep ~ lpeso+pg45,data=df))

# b
summary(step(fit, scope=list(upper = fit, lower = ~ 1)))

# c
summary(step(fit, scope=list(upper = fit, lower = ~ 1), k = log(nrow(df))))
summary(lm(laep ~ lvolca+lpeso, data=df))
AIC(lm(laep ~ lvolca+lpeso, data=df))
BIC(lm(laep ~ lvolca+lpeso, data=df))

#d
all <- data.frame(ols_step_all_possible(fit))
arrange(all,desc(adjr))
head(arrange(all,desc(adjr)))

#----- Análise de Resíduos
shapiro.test(fit3$residuals)
gqtest(fit3)
dwtest(fit3)

a <- ggplot(fit3)+
  geom_point(aes(x=.fitted,y=studres(fit3)))+
  labs(x="Valores preditos \n A",y="Resíduos estudentizados")+
  theme_minimal()

b <- ggplot(fit3, aes(sample=studres(fit3)))+
  geom_qq()+
  geom_qq_line(color="red")+
  labs(x="Quantil teórico \n B", y="Resíduos estudentizados")+
  theme_minimal()

c <- ggplot(fit3)+
  geom_histogram(aes(x=.resid),binwidth=0.3,color="black", fill="white")+
  labs(y="Frequência",x="Resíduos \n C")+
  theme_minimal()

a | b | c

ggplot(fit3)+

```

```
geom_point(aes(x=seq(1,97),y=.resid))+  
labs(x="Ordem de Coleta",y="Resíduos")+  
theme_minimal()
```