Códigos utilizados no laboratório 9

Camille Menezes e Michel Miler

Junho de 2023

```r
library(tidyverse)
library(GGally)
library(corrplot)
library(patchwork)
library(olsrr)
library(gridExtra)
library(gmodels)
library(qqplotr)
library(dplyr)


df <- read.csv("dados_mundiais.csv")[,-1]

df <- df|>
  mutate(Continente = as.factor(Continente))
df$Continente = relevel(df$Continente,"Europa")
df <- df[,-length(df)]

cor(df$ExpectHomens,log(df$RendaMensal))

ggplot(df)+
  aes(x=ExpectHomens,y=log(RendaMensal),color=Continente)+
  geom_smooth(method = lm, se = FALSE)+
  scale_x_continuous("Expectativa de vida masculina")+
  scale_y_continuous("Logaritmo da renda mensal")+
  geom_point()+
  theme_minimal()

a <- ggplot(df)+
  aes(x=Continente,y=log(RendaMensal),fill=Continente)+
  geom_boxplot()+
  scale_x_discrete("",breaks = NULL) +
  scale_y_continuous("Logaritmo da renda mensal")+
  theme_bw()+
  theme(legend.position="none")

b <- ggplot(df)+
  aes(x=Continente,y=ExpectHomens,fill=Continente)+
  geom_boxplot()+
  scale_x_discrete("",breaks = NULL) +
  scale_y_continuous("Expectativa de vida masculina")+
  theme_bw()

a | b
```

```r
#modelos
fit1 <- lm(log(RendaMensal)~1, data = df)
fit2 <- lm(log(RendaMensal)~ExpectHomens, data = df)
fit3 <- lm(log(RendaMensal)~ExpectHomens+Continente, data = df)
fit4 <- lm(log(RendaMensal)~ExpectHomens*Continente, data = df)

exp(fit1$coefficients)
summary(fit1)
summary(fit2)
summary(fit3)

anova(fit3 ,fit2)

summary(fit4)

anova(fit4,fit1)
anova(fit4,fit2)
anova(fit4,fit3)

model <- fit4


#Analise dos resíduos
shapiro.test(model$residuals)
gqtest(model)
dwtest(model)

a <- ggplot(model)+
  geom_point(aes(x=.fitted,y=studres(model),color=df$Continente))+
  scale_color_manual(values=c("#ff6961","#77dd77","#cbcd69","#84b6f4","#cd9cb2"))+
  geom_hline(yintercept = c(2,-2), colour = "black",linetype = "dashed")+
  geom_hline(yintercept = c(0), colour = "red",linetype = "dashed")+
  scale_color_discrete("Continente")+
  labs(x="Valores preditos \n B",y="Resíduo Studentizado")+
  theme_bw()


b <- ggplot(model, aes(sample=studres(model)))+
  geom_qq_line(color="black")+
  #geom_qq_band(fill="lightblue")+
  geom_qq(aes(color=df$Continente))+
  scale_color_manual("Continente",breaks=NULL,values=c("#ff6961","#77dd77","#cbcd69",
                                                       "#84b6f4","#cd9cb2"))+
  labs(x="Quantis teóricos \n A", y="Resíduo Studentizado")+
  theme_bw()


b | a
```

```r
m_infl <- influence.measures(model)
metricas <- as.data.frame(m_infl$infmat)
names(metricas)

#leverage
g <- ols_prep_rstudlev_data(model)
d <- g$levrstud
d$txt <- ifelse(d$color == "normal", NA, d$obs)
f <- d[d$color == "outlier", c("obs", "leverage", "rstudent")]
colnames(f) <- c("observation", "leverage", "stud_resid")
a <- ggplot(d, aes(leverage, rstudent, label = txt)) +
  geom_point(aes(colour = fct_color)) +
  scale_color_manual(labels = c("normal","ponto de alavanca","outlier",
                                "outlier e ponto de alavanca"),
                     values = c("black","blue", "red", "green")) +
  xlim(g$minx, g$maxx) +
  ylim(g$miny, g$maxy) +
  labs(colour = "Observação", x = "Leverage \n B", y = "Resíduo Studentizado") +
  geom_hline(yintercept = c(2,-2), colour = "black",linetype = "dashed") +
  geom_vline(xintercept = g$lev_thrsh, colour = "black",linetype = "dashed") +
  geom_text(vjust = -1, size = 3, colour = "black") +
  theme_bw()




#dist^ancia de Cook
k <- ols_prep_cdplot_data(model)
d <- ols_prep_outlier_obs(k)
f <- ols_prep_cdplot_outliers(k)
b <- ggplot(d, aes(x = obs, y = cd, label = txt)) +
  geom_bar(width = 0.2, stat = "identity",
           aes(fill = fct_color)) +
  scale_fill_manual(values = c("black", "black"),breaks=NULL) +
  ylim(0, k$maxx) +
  labs(x = "Observação \n A", y = "Distância de Cook") +
  geom_hline(yintercept = 0, colour = "white") +
  geom_hline(yintercept = k$ts, colour = "red",linetype = "dashed") +
  geom_text(hjust = -0.2, nudge_x = 0.05, size = 3, na.rm = TRUE, color="red") +
  theme_bw()

b | a
```