Lista 04

Matheus Cougias

16/01/2021

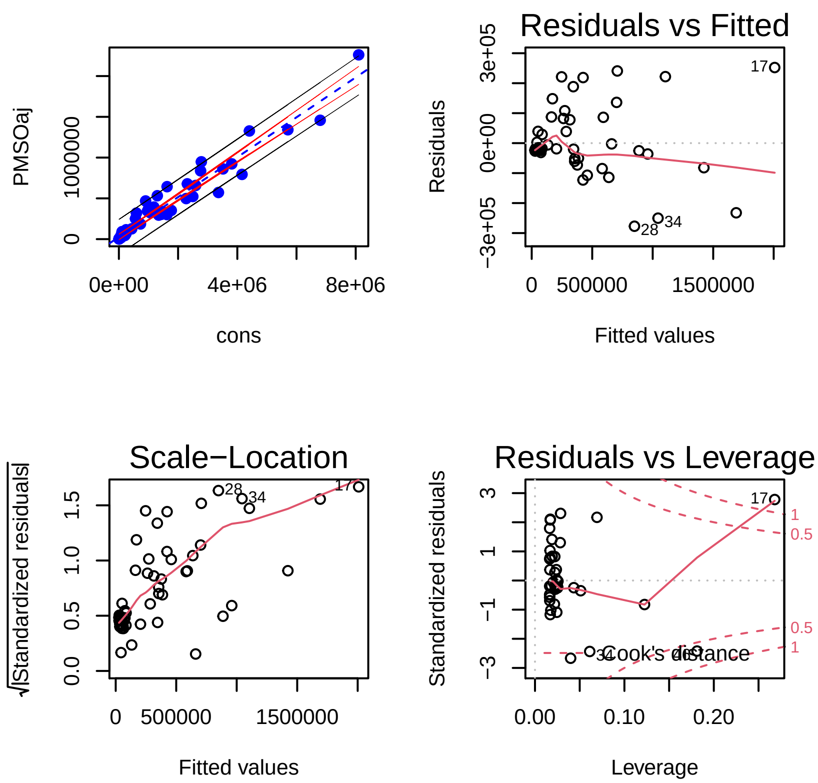
```
dados <- read.csv2("aneel_2014-2016.csv")
dadosA <- data.frame(dados$PMSOaj, dados$cons)
dadosB <- data.frame(log(dados$PMSOaj), log(dados$cons))
names(dadosA)[1] <- "PMSOaj"
names(dadosA)[2] <- "cons"
names(dadosB)[1] <- "PMSOaj"
names(dadosB)[2] <- "cons"

par(mfrow = c(2, 3))
#Modelo A
plot(PMSOaj ~ cons, data=dadosA, pch=19, col="blue")
modeloA <- lm(PMSOaj ~ cons, data=dadosA)
abline(modeloA, lty=2, col="blue", lwd=1)
summary(modeloA)
```

```
##
## Call:
## lm(formula = PMSOaj ~ cons, data = dadosA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -277235  -25469  -21440   28993  252235
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.623e+04  1.686e+04   1.556    0.125
## cons        2.447e-01  7.802e-03  31.360   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 106100 on 59 degrees of freedom
## Multiple R-squared:  0.9434, Adjusted R-squared:  0.9424
## F-statistic: 983.4 on 1 and 59 DF,  p-value: < 2.2e-16
```

```
saida <- predict(modeloA, interval = "confidence", level=0.95)
lines(saida[,"lwr"] ~ dadosA$cons, lwd=0.1, col="red")
lines(saida[,"upr"] ~ dadosA$cons, lwd=0.1, col="red")
saida <- predict(modeloA, interval = "prediction", level=0.95)
```

```
## Warning in predict.lm(modeloA, interval = "prediction", level = 0.95): predictions on current data r
```

```r
lines(saida[,"lwr"] ~ dadosA$cons, lwd=0.1, col="black")
lines(saida[,"upr"] ~ dadosA$cons, lwd=0.1, col="black")
plot(modeloA)
```




```r
par(mfrow = c(2, 3))
#Modelo B
plot(PMSOaj ~ cons, data=dadosB, pch=19, col="blue")
modeloB <- lm(PMSOaj ~ cons, data=dadosB)
abline(modeloB, lty=2, col="blue", lwd=1)
summary(modeloB)
```

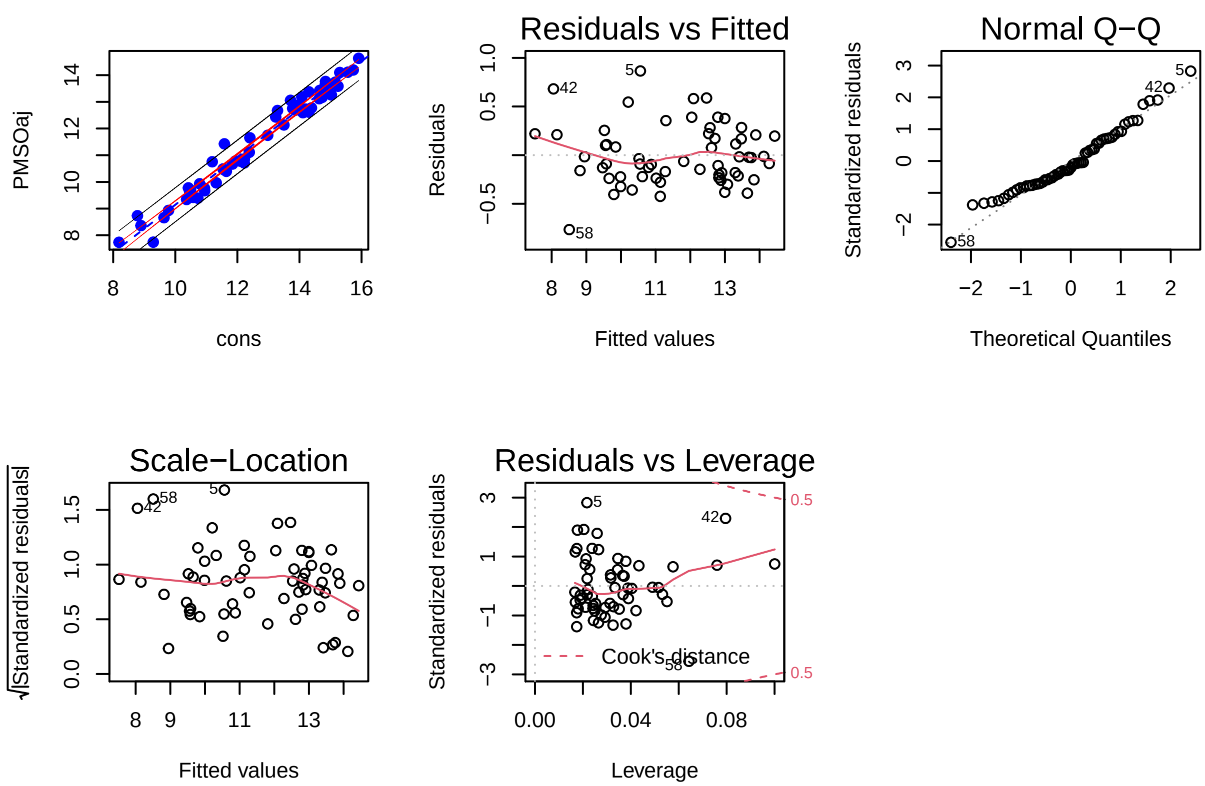
```
##
## Call:
## lm(formula = PMSOaj ~ cons, data = dadosB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76564 -0.22109 -0.06437  0.20798  0.86541
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.18413    0.25432   0.724    0.472
## cons         0.89588    0.01974  45.374   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.3096 on 59 degrees of freedom
## Multiple R-squared:  0.9721, Adjusted R-squared:  0.9717
## F-statistic:  2059 on 1 and 59 DF,  p-value: < 2.2e-16
```

```
saida <- predict(modeloB, interval = "confidence", level=0.95)
lines(saida[,"lwr"] ~ dadosB$cons, lwd=0.1, col="red")
lines(saida[,"upr"] ~ dadosB$cons, lwd=0.1, col="red")
saida <- predict(modeloB, interval = "prediction", level=0.95)
```

```
## Warning in predict.lm(modeloB, interval = "prediction", level = 0.95): predictions on current data r
```

```
lines(saida[,"lwr"] ~ dadosB$cons, lwd=0.1, col="black")
lines(saida[,"upr"] ~ dadosB$cons, lwd=0.1, col="black")
plot(modeloB)
```



Letra A

i) Com a base de dados previamente carregada, busca-se então calcular o coeficiente linear entre das variáveis PMSOaj (custo operacional) e cons (número de consumidores). Dessa maneira, ao realizar a regressão linear, chegamos que ao valor base da regressão é de B0 = 26230 e que acada quantidade acrescida de consumidores, seu valor é acrescido em B1 = 0.2447.

ii) O modelo de regressão possui tanto o R^2 múltiplo quanto o R^2 ajustado de 0.9424, o que significa que ele consegue explicar 94.24% da dispersão dos dados em relação ao modelo de média. Ao observar esse valor, pode-se dizer que o modelo compreende bem os dados, então existe sim uma certa linearidade entre os valores de PMSOaj e cons.

iii) Através da análise do gráfico de dispersão dos resíduos e de sua normal, pode-se perceber que alguns dos dados do modelo são bem compatíveis com o banco de dados original. Por outro lado, principalmente no gráfico da Normal Q-Q, é possível observar uma grande dispersão dos dados em ambas as extremidades do gráfico, com algumas observações bem acima ou abaixo do valor esperado da normal.

iv) Devido a essa grande dispersão de dados no gráfico dos resíduos, pode-se concluir que os dados não são tão compatíveis assim ao modelo ajustado em relação ao que se esperava ao analisar somente o R^2 do modelo.

Letra B

i) Utilizando os valores logaritmicos da base de dados, chega-se em B0 = 0.18413 e B1 = 0.89588.

ii) O modelo de regressão possui um R^2 múltiplo de 0.9721 e um R^2 ajustado de 0.9717, o que mostra que ele consegue explicar mais de 97% da dispersão dos dados, comprovando assim que existe uma linearidade entre os valores de Log(PMSOaj) e Log(cons).

iii) No caso do modelo onde os logaritmos são utilizados, pode-se perceber uma maior linearidade dos valores do gráfico Normal Q-Q, onde quase todos os pontos da distribuição estão extremamente próximos da reta normal montada.

iv) É possível concluir que nesse segundo modelo existe uma certa compatibilidade dos dados em relação ao banco de dados original, muito por causa de uma boa dispersão de resíduos gerada.

Letra C

Ao comparar as conclusões retiradas de ambos os modelos, percebe-se que o modelo onde o logaritmo é aplicado existe uma melhor compreensão dos dados em relação ao modelo inicial. Através de algumas comparações, podemos chegar a essa conclusão:

1 -> Se compararmos a quantidade de pontos dentro dos limites dos gráficos de regressão, é claramente visível uma maior quantidade de dados dentro dos limites do modelo logaritmico do que no modelo original;

2 -> O modelo logaritmico consegue compreender, através do valor de R^2, mais de 97% dos dados da base de dados, enquanto o modelo inicial tinha um valor de R^2 relativamente menor, de 94%.

3 -> A principal diferença percebida entre os modelos está nos gráficos dos resíduos gerados. Ao comparar o gráfico da Normal Q-Q, pode-se ver que os pontos do modelo onde o logaritmo é aplicado estão bem mais próximos da reta normal que no modelo com os dados originais.