

Lista 04-05

Matheus Cougias

30/01/2021

#Informação Importante Durante a leitura dos dados e transformação da coluna “ralta” em $\log(\text{ralta})$ houve um erro, pois alguns valores da coluna apresentaram valor zero, dessa maneira, fiz a alteração desses valores para 1, corrigindo o erro gerado e não alterando significativamente a base de dados.

```
dados <- read.csv2("aneel_2014-2016.csv")
for(i in 1:61){
  if (dados$ralta[i] == 0){
    dados$ralta[i] <- 1
  }
}
```

#Modelo Log

```
reg_multipla_log = lm(log(PMS0aj) ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons), data = dados)
summary(reg_multipla_log)
```

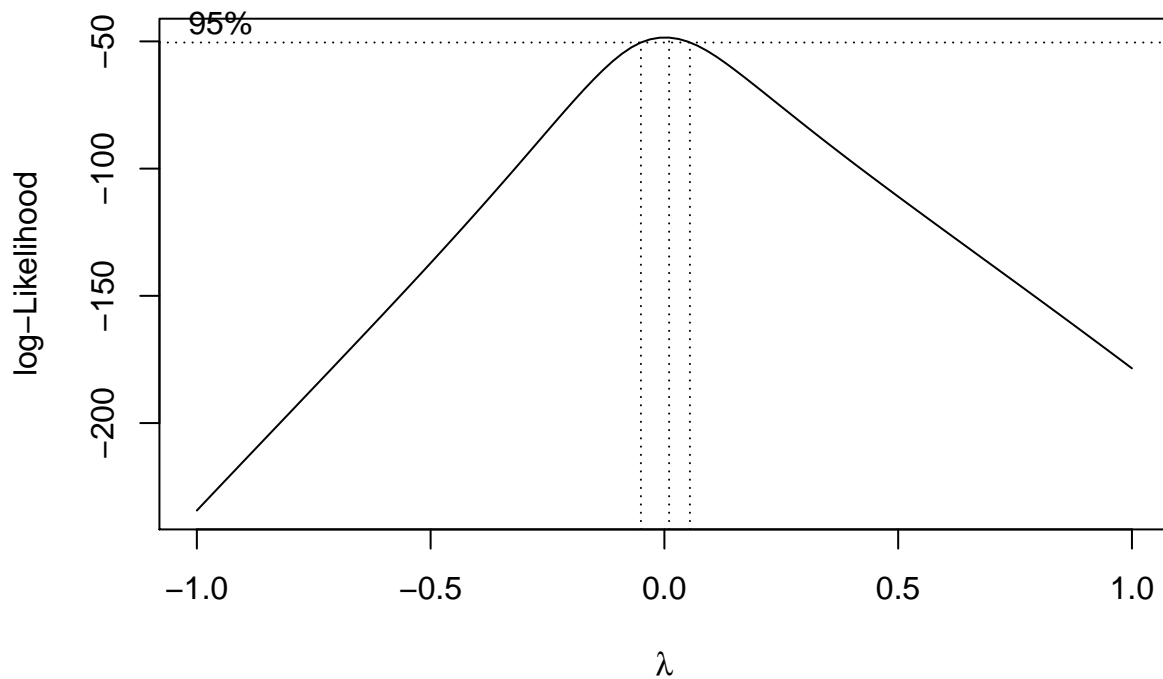
```
##
## Call:
## lm(formula = log(PMS0aj) ~ rsub + log(rdist_a) + log(ralta) +
##     log(mponderado) + log(cons), data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6614 -0.2125 -0.0086  0.1826  0.7311
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.393e-01  5.110e-01   0.664  0.50954
## rsub          4.532e-05  9.008e-05   0.503  0.61685
## log(rdist_a)  8.914e-02  6.270e-02   1.422  0.16074
## log(ralta)    2.903e-02  2.378e-02   1.221  0.22727
## log(mponderado) 3.331e-01  1.517e-01   2.196  0.03229 *
## log(cons)     4.464e-01  1.609e-01   2.775  0.00753 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2977 on 55 degrees of freedom
## Multiple R-squared:  0.976, Adjusted R-squared:  0.9738
## F-statistic: 447.1 on 5 and 55 DF, p-value: < 2.2e-16
```

#Modelo Boxcox

```
require(MASS)
```

```
## Loading required package: MASS
```

```
testePMS0aj <- boxcox(PMS0aj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons), data = d
```



```
lambda <- testePMS0aj$x[which.max(testePMS0aj$y)]
dados$boxcoxPMS0aj <- (dados$PMS0aj^lambda - 1)/lambda
reg_multipla_boxcox = lm(boxcoxPMS0aj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons),
summary(reg_multipla_boxcox)
```

```
##
```

```
## Call:
```

```
## lm(formula = boxcoxPMS0aj ~ rsub + log(rdist_a) + log(ralta) +
##     log(mponderado) + log(cons), data = dados)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.69538 -0.25747 -0.01044  0.21467  0.80043
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.128e-01  5.750e-01  -0.544  0.58871
## rsub           5.831e-05  1.014e-04   0.575  0.56746
```

```
## log(rdist_a)      9.527e-02  7.055e-02   1.350  0.18247
## log(ralta)       3.179e-02  2.675e-02   1.188  0.23991
## log(mponderado)  3.729e-01  1.707e-01   2.185  0.03316 *
## log(cons)        5.058e-01  1.810e-01   2.794  0.00715 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.335 on 55 degrees of freedom
## Multiple R-squared:  0.9758, Adjusted R-squared:  0.9736
## F-statistic: 443.4 on 5 and 55 DF,  p-value: < 2.2e-16
```

#Resultados O valor selecionado para o lambda durante a utilização do boxcox foi de 0.010101..., bem próximo do valor padrão para aplicar o logaritmo. A primeira comparação que pode ser feita é em relação aos valores de R^2 múltiplo e ajustado de cada um dos modelos. Os valores obtidos na regressão linear múltipla com a variável log(PMSOaj) foram de 97.6% e 97.38%, enquanto os valores obtidos na utilização do boxcox de PMSOaj foram de 97.58% e 97.36%, bem próximos também dos valores do logaritmo.

Esses valores próximos são explicados devido ao lambda gerado durante o boxcox, de 0.01010101, bem próximo de 0, que seria considerado como a aplicação direta do logaritmo na variável. Dessa maneira, ambos os modelos possuem resultados muito similares, podendo ambos serem utilizados.