

Lista 06

Matheus Cougias

06/02/2021

##Leitura dos dados

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

##Modelos de regressão Como não estava especificado na lista, preferi utilizar o modelo baseado no PMSOaj do que o modelo baseado no log(PMSOaj). Outro ponto não especificado foi dado pelos modelos montados combinando as variáveis, onde não foram especificadas as combinações a serem utilizadas. Dessa maneira, montei as regressões lineares simples com cada uma das variáveis e, as combinações preferi gerar adicionando as variáveis em ordem de acordo com o descrito no arquivo.

```
modelo1 <- lm(PMSOaj ~ rsub, data = dados)
modelo2 <- lm(PMSOaj ~ log(rdist_a), data = dados)
modelo3 <- lm(PMSOaj ~ log(ralta), data = dados)
modelo4 <- lm(PMSOaj ~ log(mponderado), data = dados)
modelo5 <- lm(PMSOaj ~ log(cons), data = dados)
modelo6 <- lm(PMSOaj ~ PNTaj, data = dados)
modelo7 <- lm(PMSOaj ~ CHIaj, data = dados)
modelo8 <- lm(PMSOaj ~ rsub + log(rdist_a), data = dados)
modelo9 <- lm(PMSOaj ~ rsub + log(rdist_a) + log(ralta), data = dados)
modelo10 <- lm(PMSOaj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado), data = dados)
modelo11 <- lm(PMSOaj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons), data = dados)
modelo12 <- lm(PMSOaj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons) + PNTaj, data = dados)
modelo13 <- lm(PMSOaj ~ rsub + log(rdist_a) + log(ralta) + log(mponderado) + log(cons) + PNTaj + CHIaj,
```

##Análise dos coeficientes de determinação preditivo A partir dos modelos de regressão gerados, busca-se analisar os resultados apresentados pelos coeficientes baseando na matriz de projeção H de cada um dos modelos. Em relação ao modelo onde todas as variáveis foram utilizadas, o seguinte erro surgiu: sistema é computacionalmente singular, fazendo com que fosse excluído do estudo. Assim, ao analisar o coeficiente de cada um dos modelos montados, o que apresentou o melhor resultado foi o modelo 6, que relaciona a variável PMSOaj com a variável preditora PNTaj, onde seu coeficiente de determinação preditivo foi de 78,6551%.

```
Y <- as.matrix(dados$PMSOaj)
SQT <- sum((dados$PMSOaj - mean(dados$PMSOaj))^2)

X1 <- model.matrix(modelo1)
H1 <- X1%*%solve(t(X1)%*%X1)%*%t(X1)
```

```

res1 <- (Y - H1%*%Y)
SQE1 <- sum(res1^2)
coef1 <- SQE1/SQT

X2 <- model.matrix(modelo2)
H2 <- X2%*%solve(t(X2)%*%X2)%*%t(X2)
res2 <- (Y - H2%*%Y)
SQE2 <- sum(res2^2)
coef2 <- SQE2/SQT

X3 <- model.matrix(modelo3)
H3 <- X3%*%solve(t(X3)%*%X3)%*%t(X3)
res3 <- (Y - H3%*%Y)
SQE3 <- sum(res3^2)
coef3 <- SQE3/SQT

X4 <- model.matrix(modelo4)
H4 <- X4%*%solve(t(X4)%*%X4)%*%t(X4)
res4 <- (Y - H4%*%Y)
SQE4 <- sum(res4^2)
coef4 <- SQE4/SQT

X5 <- model.matrix(modelo5)
H5 <- X5%*%solve(t(X5)%*%X5)%*%t(X5)
res5 <- (Y - H5%*%Y)
SQE5 <- sum(res5^2)
coef5 <- SQE5/SQT

X6 <- model.matrix(modelo6)
H6 <- X6%*%solve(t(X6)%*%X6)%*%t(X6)
res6 <- (Y - H6%*%Y)
SQE6 <- sum(res6^2)
coef6 <- SQE6/SQT

X7 <- model.matrix(modelo7)
H7 <- X7%*%solve(t(X7)%*%X7)%*%t(X7)
res7 <- (Y - H7%*%Y)
SQE7 <- sum(res7^2)
coef7 <- SQE7/SQT

X8 <- model.matrix(modelo8)
H8 <- X8%*%solve(t(X8)%*%X8)%*%t(X8)
res8 <- (Y - H8%*%Y)
SQE8 <- sum(res8^2)
coef8 <- SQE8/SQT

X9 <- model.matrix(modelo9)
H9 <- X9%*%solve(t(X9)%*%X9)%*%t(X9)
res9 <- (Y - H9%*%Y)
SQE9 <- sum(res9^2)
coef9 <- SQE9/SQT

X10 <- model.matrix(modelo10)

```

```

H10 <- X10%*%solve(t(X10)%*%X10)%*%t(X10)
res10 <- (Y - H10%*%Y)
SQE10 <- sum(res10^2)
coef10 <- SQE10/SQT

X11 <- model.matrix(modelo11)
H11 <- X11%*%solve(t(X11)%*%X11)%*%t(X11)
res11 <- (Y - H11%*%Y)
SQE11 <- sum(res11^2)
coef11 <- SQE11/SQT

X12 <- model.matrix(modelo12)
H12 <- X12%*%solve(t(X12)%*%X12)%*%t(X12)
res12 <- (Y - H12%*%Y)
SQE12 <- sum(res12^2)
coef12 <- SQE12/SQT

```

##Análise dos resíduos Através da análise dos resíduos é possível perceber uma baixa quantidade de observações discrepantes. A observação 17 apresentou um resíduo padronizado de 3,756, a observação 49 apresentou um resíduo padronizado de 2,972 e a observação 2 apresentou de -2,952. Já, ao utilizarmos o resíduo deletado, podemos perceber que a predição do modelo encontra-se em faixa bem alta, próxima de 1.

```

dados <- subset(dados, select = c("PMS0aj", "PNTaj"))

Beta = solve(t(X6)%*%X6)%*%t(X6)%*%Y
I <- diag(rep(1,nrow(dados)))
sigma2 <- t(res6)%*%res6/(length(Y) - ncol(X6))
sigma <- as.numeric(sqrt(sigma2))
residuosPadronizados <- res6/sqrt(sigma2*diag(I - H6))

```

```

## Warning in sigma2 * diag(I - H6): Recycling array of length 1 in array-vector arithmetic is deprecated
## Use c() or as.vector() instead.

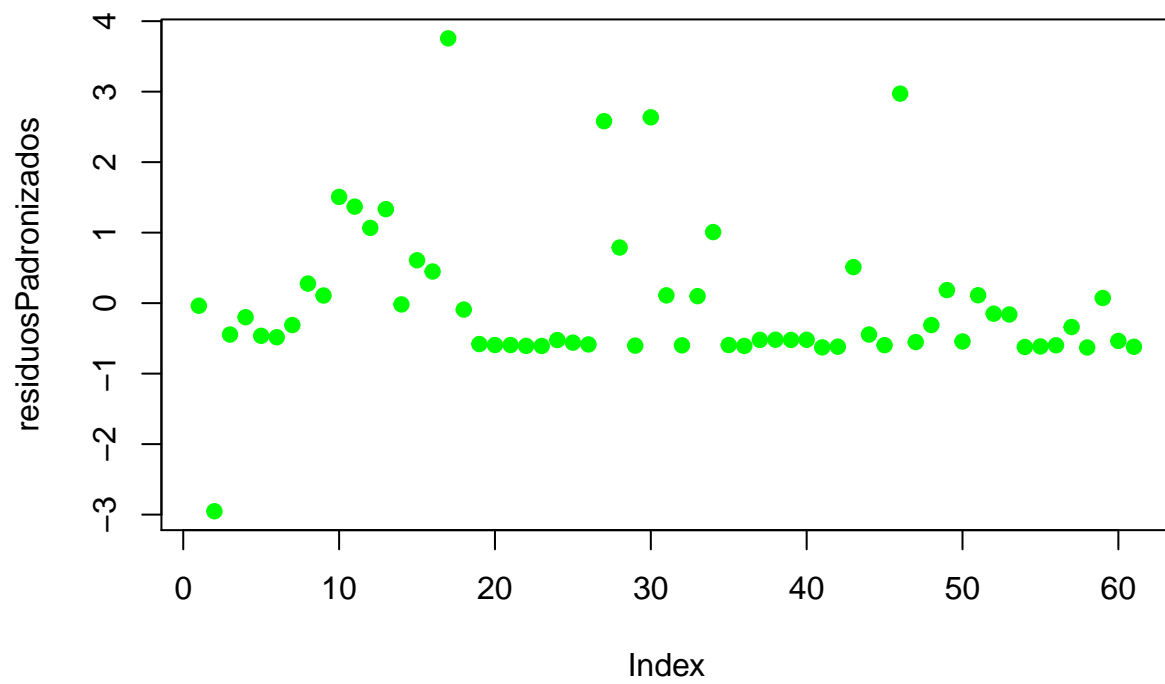
```

```

dados$Padronizados <- residuosPadronizados

plot(residuosPadronizados, col="green", pch=19)

```



```
resDeletado <- res6/as.numeric(diag(I-H6))  
press <- sum(resDeletado)  
R2Pred <- 1 - press/SQT
```