

Lista 07

Matheus Cougias e Klysman Rezende

14/09/2020

Leitura da base de dados

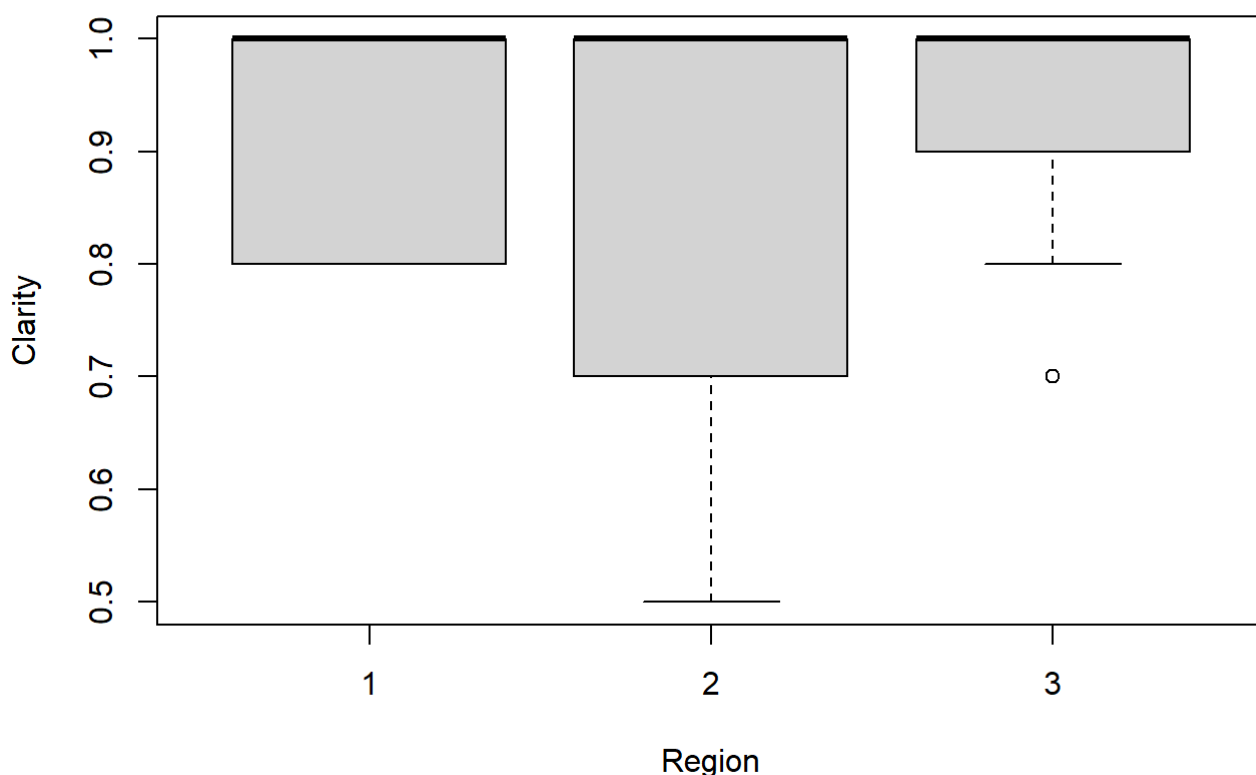
Realiza a leitura da base de dados, que possui 38 observações e 7 variáveis, sendo a Region a variável a ser analisada. As variáveis Sabor e Qualidade serão desconsideradas na análise.

```
dados <- read.csv("wine.csv")
dados = subset(dados, select = -c(Flavor, Quality))
names(dados)[names(dados) == 'Region.y.'] <- 'Region'
dados$Region <- as.factor(dados$Region)
```

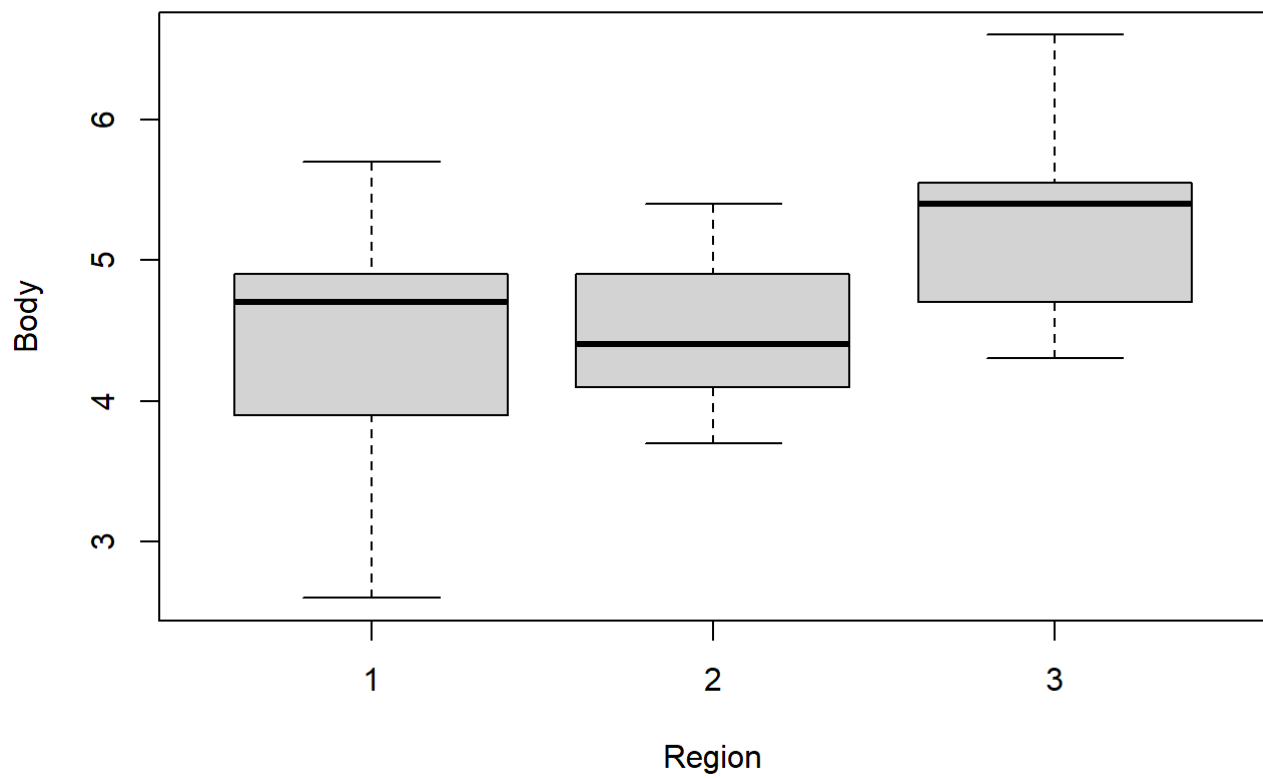
Análise exploratória

Observando a base de dados, percebe-se que existem três possíveis regiões de produção de vinho, chamadas de região: 1, 2 e 3. Através de uma análise inicial, percebe-se que a região 3 destaca-se das demais pelos valores mais altos no Aroma e no Corpo, enquanto a região 2 se diferencia das demais por uma variedade maior de Claridade.

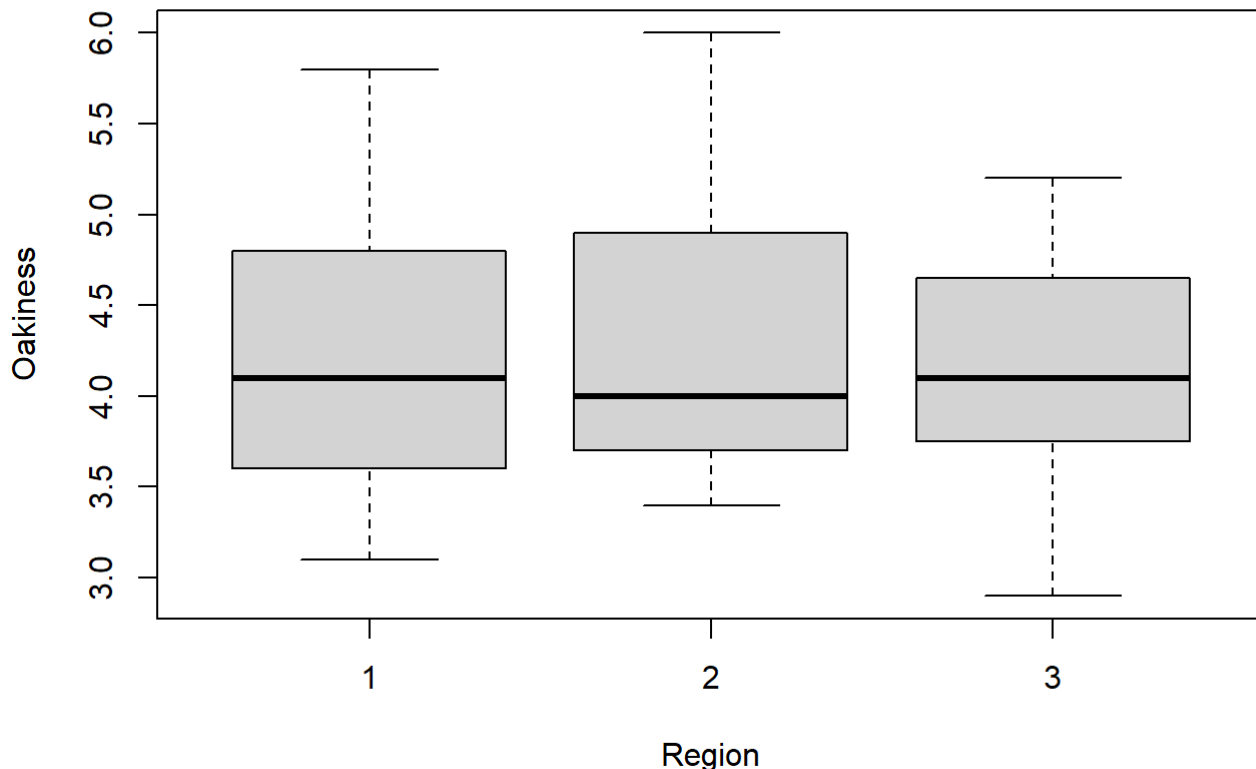
```
plot(Clarity ~ Region, data=dados)
```



```
#plot(Aroma ~ Region, data=dados)  
plot(Body ~ Region, data=dados)
```



```
plot(Oakiness ~ Region, data=dados)
```



Modelo Multinomial

Dessa maneira, inicia-se a utilização do pacote `mlogit` repassando para ele a base de dados, além de selecionar qual será a variável resposta do problema. No segundo passo, foi escolhida a região 1 como base de comparação dos resultados da regressão. Dessa maneira, pode-se observar como cada uma das características se comporta de acordo com a região em que o vinho é produzida, então: -> Claridade: Nessa característica pode-se observar a maior dissiparidade entre as regiões, onde a região 3 possui uma tendência muito acima das demais de produzir um vinho com a variável Claridade mais alta; -> Aroma: Novamente a região 3 destaca-se em mais uma variável, onde os vinhos dessa região possuem maiores valores para a variável; -> Corpo: Entre as variáveis, a mais “equilibrada” foi o Corpo do vinho, com valores extremamente próximos. Nesse caso, a região 3 possui maior possibilidade de produzir vinhos com sua variável Corpo com valores mais altos; -> Carvalho: Se diferenciando um pouco das demais variáveis, o carvalho da região 3 tem uma tendência a valores um pouco menores em relação às demais regiões. Nessa variável a região 2 tem maior possibilidade de valores mais altos.

```
require(mlogit)
```

```
## Loading required package: mlogit
```

```
## Loading required package: dfidx
```

```
##  
## Attaching package: 'dfidx'
```

```
## The following object is masked from 'package:stats':
##
##   filter
```

```
mdados <- mlogit.data(dados, choice="Region", shape="wide")
mdados.model <- mlogit(Region ~ 1|Clarity + Aroma + Body + Oakiness, data = mdados, reflevel=
1)
summary(mdados.model)
```

```
##
## Call:
## mlogit(formula = Region ~ 1 | Clarity + Aroma + Body + Oakiness,
##       data = mdados, reflevel = 1, method = "nr")
##
## Frequencies of alternatives:choice
##      1      2      3
## 0.44737 0.23684 0.31579
##
## nr method
## 9 iterations, 0h:0m:0s
## g'(-H)^-1g = 1.99E-06
## successive function values within tolerance limits
##
## Coefficients :
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept):2   3.397030   4.751775   0.7149   0.47467
## (Intercept):3 -48.245687  26.181983  -1.8427   0.06537 .
## Clarity:2      -4.887302   3.702180  -1.3201   0.18680
## Clarity:3      21.765467  15.931310   1.3662   0.17187
## Aroma:2       -0.088522   0.734457  -0.1205   0.90407
## Aroma:3        5.702215   3.288971   1.7337   0.08296 .
## Body:2       -0.160094   0.657092  -0.2436   0.80751
## Body:3        3.628352   2.485322   1.4599   0.14431
## Oakiness:2     0.347561   0.618277   0.5621   0.57402
## Oakiness:3    -4.912521   2.756899  -1.7819   0.07477 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -20.951
## McFadden R^2:  0.4823
## Likelihood ratio test : chisq = 39.037 (p.value = 4.8375e-06)
```

Previsão

Somente para fim de testes, tomamos alguns valores médios, altos e baixos, respectivamente, para gerar as probabilidades das regiões terem produzido o vinho com as seguintes características (dentro dos limites encontrados na Análise Exploratória): -> Claridade 0.75, Aroma 6, Corpo 4.5, Carvalho 4.5 = Região 1 (48.17%), Região 2 (50.34%), Região 3 (1.47%) -> Claridade 0.5, Aroma 4, Corpo 3, Carvalho 3 = Região 1 (23.85%), Região 2 (76.16%), Região 3 (0.00%) -> Claridade 1, Aroma 6, Corpo 4.5, Carvalho 4.5 = Região 1 (0.32%), Região 2 (0.12%), Região 3 (99.55%)

```
teste1 <- predict(mdados.model,  
                  newdata=data.frame(Clarity=rep(0.75, 3),  
                                     Aroma=rep(6, 3),  
                                     Body=rep(4.5, 3),  
                                     Oakiness=rep(4.5, 3)))  
  
print(teste1)
```

```
##           1           2           3  
## 0.48178742 0.50347429 0.01473829
```

```
teste2 <- predict(mdados.model,  
                  newdata=data.frame(Clarity=rep(0.5, 3),  
                                     Aroma=rep(4, 3),  
                                     Body=rep(3, 3),  
                                     Oakiness=rep(3, 3)))  
  
print(teste2)
```

```
##           1           2           3  
## 2.383592e-01 7.616408e-01 2.417396e-09
```

```
teste3 <- predict(mdados.model,  
                  newdata=data.frame(Clarity=rep(1, 3),  
                                     Aroma=rep(7, 3),  
                                     Body=rep(6, 3),  
                                     Oakiness=rep(6, 3)))  
  
print(teste3)
```

```
##           1           2           3  
## 0.003231777 0.001206733 0.995561490
```