

## MAS: Trabalho de Grupo

Maria João Ferreira Lourenço

21 de março, 2023

Preencher a identificação do grupo:

**NÚMERO DO GRUPO: 7 LISTA DE TODOS OS ELEMENTOS DO GRUPO (Número - nome):**

Eliane Gabriel Nº 103303 Marco Esperança Nº 110451 Maria João Lourenço Nº104716  
Umeima Mahomed Nº 99239

O Trabalho de Grupo de *Métodos de Aprendizagem Supervisionada* refere-se à análise do data set “Consumo.Jovens.csv”.

Neste data set incluem-se 1523 registos e 28 atributos listados a seguir:

**q0:** País de residência

**q1:** Sexo

**q2:** Idade

**q3:** Situação estudantil

**q10:** Compra produtos de marca? (1-Sim; 2-Não)

**q12b\_a:** Compra em centros comerciais? (1-Sim; 0-Não)

**q12b\_b:** Compra em super/hipermercados? (1-Sim; 0-Não)

**q12b\_c:** Compra no comércio local? (1-Sim; 0-Não)

**q13a:** Fidelidade a marcas? (1-Sim; 0-Não)

**q13b:** Fidelidade a lojas? (1-Sim; 0-Não)

Variáveis q14 na Escala 1-Nada Importante, 2, 3, 4, 5-Extremamente importante)

**q14a:** Preço

**q14b:** Necessidade do produto

**q14c:** Conveniência da localização da loja

**q14d:** Qualidade do produto

**q14e:** Imagem do produto

**q14f:** Imagem da loja

**q14g:** Características do produto

**q14h:** Promoção especial

**q14i:** Imagem da marca

**q14j:** Publicidade

Variáveis q19 na Escala 1-Discordo Completamente, 2, 3, 4, 5-Concordo Completamente)

**q19\_1:** Alguns dos feitos + importantes da vida incluem adquirir bens materiais

**q19\_2:** Não dou importância à quantidade de bens materiais

**q19\_3:** Gosto de ter coisas para impressionar as pessoas

**q19\_4:** Geralmente compro apenas aquilo de que preciso

q19\_5: Gosto de gastar dinheiro em coisas que não são necessárias

q19\_6: Comprar coisas dá-me imenso prazer

q19\_7: Tenho todas as coisas de que preciso para ser feliz

q19\_8: Seria mais feliz se tivesse dinheiro para comprar mais coisas

## Notas:

1. Efetuar todos os Save com “Save with encoding UTF-8” de modo a manter palavras acentuadas e caracteres especiais\*\*
2. A cotação está anexa a cada pergunta
3. **OS ALUNOS QUE NÃO SUBMETEREM PDF NO MOODLE TERÃO UMA PENALIZAÇÃO DE 1 VALOR; SE, O FICHEIRO ALTERNATIVO QUE SUBMETEREM (VIA EMAIL) REPORTAR ERROS NA COMPILAÇÃO, TERÃO UMA PENALIZAÇÃO ADICIONAL DE 1 VALOR**

```
# Remover tudo!
rm(list=ls(all=TRUE))#
# Incluir as libraries de que necessita
library(MASS) # dataset's da base de dados do R
library(Metrics) # calculo de métricas
library(caret) # folds para validação cruzada; métricas de classificação

## Loading required package: ggplot2

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following objects are masked from 'package:Metrics':
##
##   precision, recall

library(lsr) # Cramer's V , eta, medida de associação
library(e1071) # Naive Bayes
library(psych) # estatística descritiva

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##   %+%, alpha

library(ggplot2) # gráficos
library(tree) # árvores de decisão
library(nnet) # regressao logistica multinominal
library(car) # multicolineariedade

## Loading required package: carData
```

```
##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##      logit

library(FNN) # KNN
library(knitr)
library(tidyverse)

## — Attaching core tidyverse packages — tidyverse
2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats   1.0.0      ✓ stringr   1.5.1
## ✓ lubridate 1.9.3      ✓ tibble    3.2.1
## ✓ purrr     1.0.2      ✓ tidyr     1.3.0

## — Conflicts —
tidyverse_conflicts() —
## ✗ psych::%+%( ) masks ggplot2::%+%( )
## ✗ psych::alpha( ) masks ggplot2::alpha( )
## ✗ dplyr::filter( ) masks stats::filter( )
## ✗ dplyr::lag( ) masks stats::lag( )
## ✗ purrr::lift( ) masks caret::lift( )
## ✗ dplyr::recode( ) masks car::recode( )
## ✗ dplyr::select( ) masks MASS::select( )
## ✗ purrr::some( ) masks car::some( )
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

## 1. Leitura dos dados “Consumo.Jovens.csv” e análise preliminar dos mesmos

### 1.1) [1 valor] Leitura dos dados; apresentação de dimensão e estrutura dos dados; verificação do número de casos com dados em falta (para todos os atributos); sumário dos dados completos (depois de eliminação dos casos/linhas com dados omissos )

```
#Leitura dos dados (Nota: verifique sep no ficheiro de origem)
CJ<-read.csv("Consumo.Jovens.csv", header=TRUE, dec=".", na.strings="",
sep=";", stringsAsFactors = TRUE)

# apresentação de dimensão e estrutura dos dados.
dim(CJ)


## [1] 1523    28
```

```

str(CJ)

## 'data.frame':    1523 obs. of  28 variables:
## $ q0      : Factor w/ 6 levels "Alemanha","China",...: 6 6 6 6 6 6 6 6 6 6
## ...
## $ q1      : Factor w/ 2 levels "Feminino","Masculino": 1 2 1 2 1 2 1 2 2 2
## ...
## $ q2      : int   19 19 19 20 21 19 20 21 20 22 ...
## $ q3      : Factor w/ 3 levels "Estudante a tempo inteiro",...: 1 1 1 1 1 1
1 1 1 1 ...
## $ q10     : Factor w/ 2 levels "Nao","Sim": 2 2 1 2 2 2 2 2 2 2 ...
## $ q12b_a: int    0 1 0 1 1 1 1 0 1 1 ...
## $ q12b_b: int    1 0 0 1 0 1 1 0 1 1 ...
## $ q12b_c: int    0 1 1 0 0 0 0 1 0 0 ...
## $ q13a   : int    0 1 0 1 0 0 1 1 0 1 ...
## $ q13b   : int    0 1 1 0 0 0 0 0 1 NA ...
## $ q14a   : int    5 4 5 3 5 3 3 5 3 3 ...
## $ q14b   : int    3 3 5 5 3 4 5 3 5 3 ...
## $ q14c   : int   NA 2 3 2 2 1 2 1 2 2 ...
## $ q14d   : int    4 5 5 4 5 3 5 4 3 4 ...
## $ q14e   : int   NA 3 3 2 4 2 3 2 3 3 ...
## $ q14f   : int   NA 4 4 2 3 1 3 1 2 2 ...
## $ q14g   : int   NA 5 5 3 4 4 4 3 4 2 ...
## $ q14h   : int   NA 2 2 2 2 3 4 2 3 2 ...
## $ q14i   : int    1 3 3 3 3 2 3 2 3 2 ...
## $ q14j   : int    2 3 3 2 2 2 4 1 2 2 ...
## $ q19_1  : int   NA 2 4 NA 3 1 NA 2 2 2 ...
## $ q19_2  : int   NA 4 4 NA 5 4 NA 4 2 4 ...
## $ q19_3  : int   NA 1 1 NA 2 1 NA 2 3 3 ...
## $ q19_4  : int   NA 3 5 NA 3 4 NA 5 3 3 ...
## $ q19_5  : int   NA 3 1 NA 3 1 NA 2 3 2 ...
## $ q19_6  : int   NA 3 3 NA 3 3 NA 3 3 1 ...
## $ q19_7  : int   NA 4 3 NA 5 2 NA 3 5 3 ...
## $ q19_8  : int   NA 3 5 NA 3 4 NA 4 4 4 ...

# Verificação do número de casos com dados em falta (para todos os atributos)
na_count <- CJ %>% summarise_all(funs(sum(is.na(.))))

## Warning: `funs()` was deprecated in dplyr 0.8.0.
##  Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`: tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

na_count

```

```
## q0 q1 q2 q3 q10 q12b_a q12b_b q12b_c q13a q13b q14a q14b q14c q14d q14e
q14f
## 1 0 5 0 21 44 4 5 7 60 70 13 19 24 14 20
23
```

```
## q14g q14h q14i q14j q19_1 q19_2 q19_3 q19_4 q19_5 q19_6 q19_7 q19_8
## 1 23 21 19 23 46 48 46 44 52 47 52 53
```

```
# eliminação dos casos/linhas com dados omissos
CJ<-na.omit(CJ)
```

```
# sumário dos dados completos
```

```
summary(CJ)
```

```
## q0 q1 q2
## Alemanha :113 Feminino :727 Min. :17.00
## China :170 Masculino:538 1st Qu.:20.00
## Espanha :266 Median :21.00
## Macau :156 Mean :21.19
## Mocambique:158 3rd Qu.:23.00
## Portugal :402 Max. :25.00
## q3 q10 q12b_a q12b_b
## Estudante a tempo inteiro:1044 Nao:556 Min. :0.0000 Min.
:0.0000
## Estudante-trabalhador : 116 Sim:709 1st Qu.:0.0000 1st
Qu.:0.0000
## Outra : 105 Median :1.0000 Median
:0.0000
## Mean :0.5209 Mean
:0.3621
## 3rd Qu.:1.0000 3rd
Qu.:1.0000
## Max. :1.0000 Max.
:1.0000
## q12b_c q13a q13b q14a
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :1.000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:3.000
## Median :0.0000 Median :0.0000 Median :0.0000 Median :4.000
## Mean :0.4791 Mean :0.4198 Mean :0.4806 Mean :3.696
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:4.000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :5.000
## q14b q14c q14d q14e
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:4.000 1st Qu.:2.000
## Median :4.000 Median :3.000 Median :4.000 Median :3.000
## Mean :3.704 Mean :2.553 Mean :4.029 Mean :2.952
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:5.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## q14f q14g q14h q14i
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
```

```
## 1st Qu.:2.000 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:2.000
## Median :2.000 Median :4.000 Median :3.000 Median :3.000
## Mean :2.544 Mean :3.496 Mean :2.651 Mean :2.675
## 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## q14j q19_1 q19_2 q19_3 q19_4
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000 Min.
:1.0
## 1st Qu.:2.000 1st Qu.:2.00 1st Qu.:3.000 1st Qu.:1.000 1st
Qu.:2.0
## Median :2.000 Median :3.00 Median :4.000 Median :2.000 Median
:3.0
## Mean :2.192 Mean :2.91 Mean :3.404 Mean :2.436 Mean
:3.3
## 3rd Qu.:3.000 3rd Qu.:4.00 3rd Qu.:4.000 3rd Qu.:3.000 3rd
Qu.:4.0
## Max. :5.000 Max. :5.00 Max. :5.000 Max. :5.000 Max.
:5.0
## q19_5 q19_6 q19_7 q19_8
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.00
## 1st Qu.:2.000 1st Qu.:3.00 1st Qu.:2.000 1st Qu.:3.00
## Median :2.000 Median :3.00 Median :3.000 Median :3.00
## Mean :2.387 Mean :3.27 Mean :2.947 Mean :3.24
## 3rd Qu.:3.000 3rd Qu.:4.00 3rd Qu.:4.000 3rd Qu.:4.00
## Max. :5.000 Max. :5.00 Max. :5.000 Max. :5.00
```

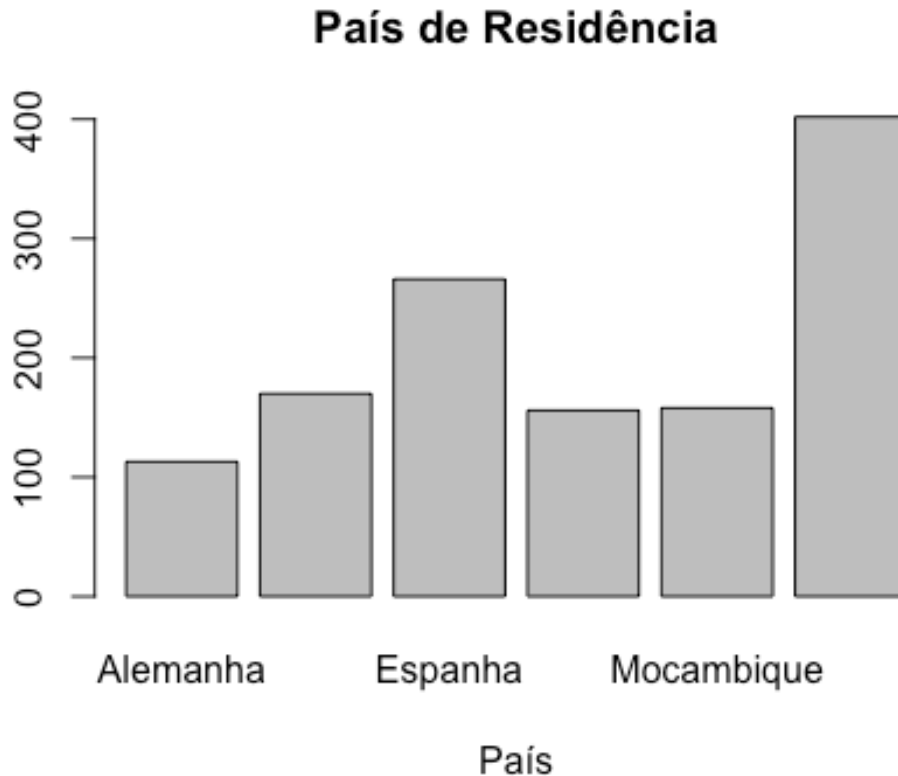
## 1.2) [1.5 valores] Breve análise descritiva de q0, q1, q2 e q3.

*#q0: País de residência*

```
summary(CJ[,1])
```

```
## Alemanha      China      Espanha      Macau Mocambique      Portugal
##          113          170          266          156          158          402
```

```
counts <- table(CJ$q0)
barplot(counts, main="País de Residência",
        xlab="País")
```



*#q1: Sexo*

```
summary(CJ[,2])
```

```
##  Feminino Masculino
```

```
##      727      538
```

```
q1_counts <- table(CJ$q1)
```

```
cores <- terrain.colors(length(q1_counts))
```

```
percentagens <- round(q1_counts / sum(q1_counts) * 100, 1)
```

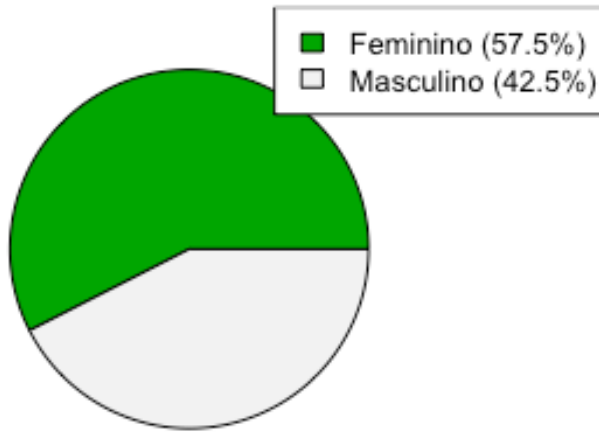
```
labels <- c("Feminino", "Masculino")
```

```
etiquetas_percentagens <- paste0(labels, " (", percentagens, "%)")
```

```
pie(percentagens, col = cores, labels = rep("", length(labels)), main =  
"Gêneros")
```

```
legend("topright", legend = etiquetas_percentagens, cex = 0.8, fill = cores)
```

## Géneros



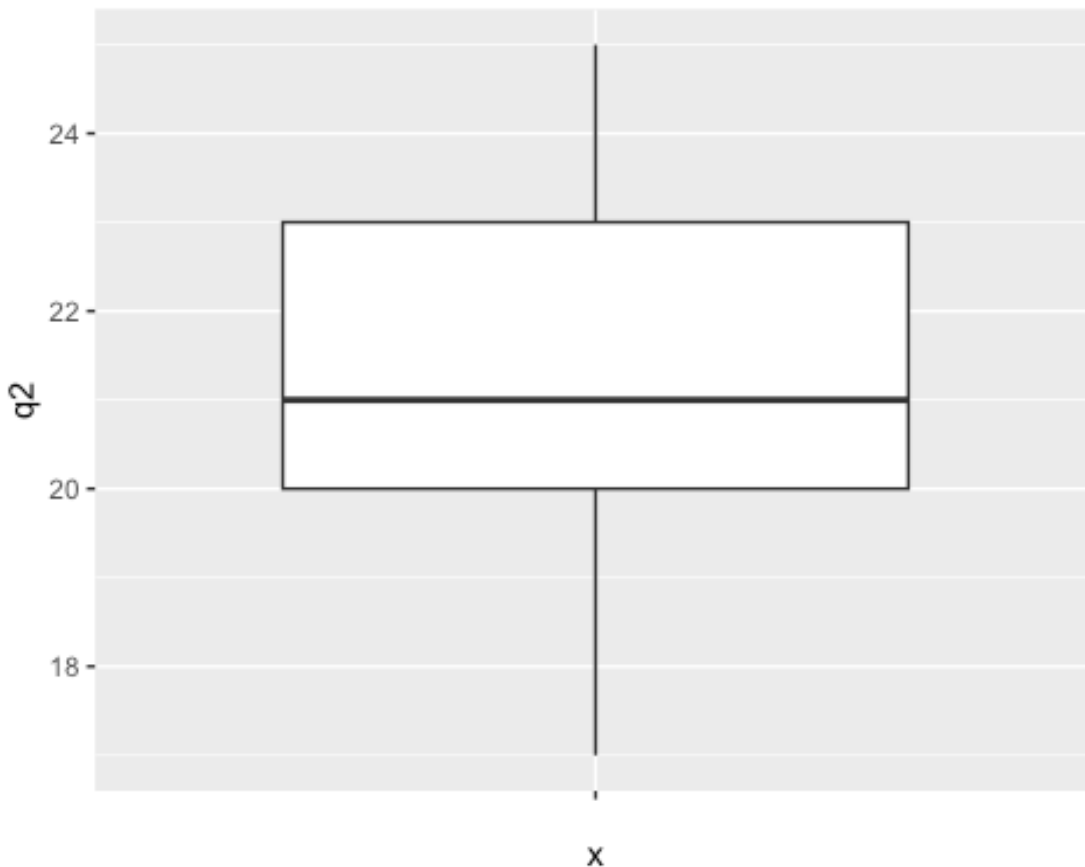
*#q2: Idade*

```
summary(CJ[,3])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.00   20.00   21.00   21.19   23.00   25.00
```

```
ggplot(CJ, aes(x = "", y = q2)) + geom_boxplot()
```





*#q3: Situação estudantil*

`summary(CJ[,4])`

```
## Estudante a tempo inteiro      Estudante-trabalhador
Outra
##                1044                116
105
```

`q3_counts <- table(CJ$q3)`

`cores <- terrain.colors(length(q3_counts))`

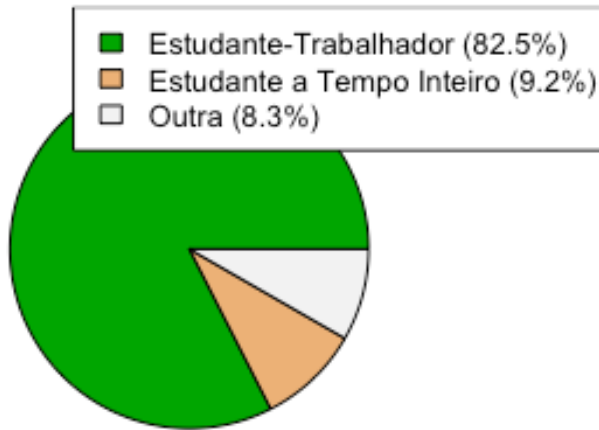
`percentagens <- round(q3_counts / sum(q3_counts) * 100, 1)`

`labels <- c("Estudante-Trabalhador", "Estudante a Tempo Inteiro", "Outra")`  
`etiquetas_percentagens <- paste0(labels, " (", percentagens, "%)")`

`pie(percentagens, col = cores, labels = rep("", length(labels)), main = "Gêneros")`

`legend("topright", legend = etiquetas_percentagens, cex = 0.8, fill = cores)`

## Gêneros



```
# todas as juntas
describe(CJ[,1:4])

##      vars      n mean   sd median trimmed  mad min max range  skew kurtosis
## q0*      1 1265  4.01 1.72      4    4.13 2.97   1  6    5 -0.23   -1.29
## q1*      2 1265  1.43 0.49      1    1.41 0.00   1  2    1  0.30   -1.91
## q2       3 1265 21.19 1.96     21   21.13 1.48  17 25    8  0.23   -0.77
## q3*      4 1265  1.26 0.60      1    1.09 0.00   1  3    2  2.16    3.23
## q0*      1 1265  4.01 1.72      4    4.13 2.97   1  6    5 -0.23   -1.29
## q1*      2 1265  1.43 0.49      1    1.41 0.00   1  2    1  0.30   -1.91
## q2       3 1265 21.19 1.96     21   21.13 1.48  17 25    8  0.23   -0.77
## q3*      4 1265  1.26 0.60      1    1.09 0.00   1  3    2  2.16    3.23
```

### 1.3) [1.5 valores] Cálculo (e apresentação) de medidas de associação entre as variáveis: a) q14a...q14j; b) q0 e as variáveis q19\_1...q19\_8; c) q10 e q1

```
Eta_ <- function(y,x){
  freqk <- as.vector(table(x))
  l <- nlevels(x)
  m <- rep(NA,l)
  qual <- as.numeric(x)
  for (k in 1:l) {m[k] <- mean(y[qual == k])}
  return(sqrt(sum(freqk*(m-mean(y))^2)/sum((y-mean(y))^2)))}
```

```

#a) q14a...q14j
# Criando uma matriz com as variáveis q14a...q14j
matrizq14 <- CJ[,11:20]

# Aplicando a função Eta para cada combinação de variáveis
medidas_assoc <- apply(matrizq14, 2, function(x) Eta_(CJ$q13a, x))

# Visualizando as medidas de associação
medidas_assoc

##           q14a           q14b           q14c           q14d           q14e           q14f
## 0.230123370 0.227084082 0.103365304 0.090701177 0.391263122 0.235840609
##           q14g           q14h           q14i           q14j
## 0.017848481 0.002613442 0.600735876 0.126387033

#b) q0 e as variáveis q19_1...q19_8

(a_q0_q19_1 <- cramersV(CJ$q0, CJ$q19_1))
## [1] 0.2079923

(a_q0_q19_2 <- cramersV(CJ$q0, CJ$q19_2))
## [1] 0.1746758

(a_q0_q19_3 <- cramersV(CJ$q0, CJ$q19_3))
## [1] 0.3678313

(a_q0_q19_4 <- cramersV(CJ$q0, CJ$q19_4))
## [1] 0.1604731

(a_q0_q19_5 <- cramersV(CJ$q0, CJ$q19_5))
## Warning in stats::chisq.test(...): Chi-squared approximation may be
incorrect
## [1] 0.1967741

(a_q0_q19_6 <- cramersV(CJ$q0, CJ$q19_6))
## [1] 0.1271594

(a_q0_q19_7 <- cramersV(CJ$q0, CJ$q19_7))
## [1] 0.242913

(a_q0_q19_8 <- cramersV(CJ$q0, CJ$q19_8))
## [1] 0.1376108

```

```
# c) q10 e q1
```

```
cramersV(CJ$q10,CJ$q1)
```

```
## [1] 0.07398348
```

**1.4) [1 valor] Divisão dos dados em amostra de treino (60%) - CJ.train - e de teste (40%) – CJ.test - usando set.seed(444);apresentação de tabela de frequências relativas de q1 em cada amostra**

```
set.seed(444)
```

```
#CJ.train
```

```
C_train <- sample(nrow(CJ),.60*nrow(CJ))
```

```
CJ.train <- CJ[C_train,]
```

```
dim(CJ.train)
```

```
## [1] 759 28
```

```
#CJ.test
```

```
CJ.test <- CJ[-C_train,]
```

```
dim(CJ.test)
```

```
## [1] 506 28
```

```
CJ2 <- table(CJ.test$q1)
```

```
knitr::kable(prop.table(CJ2))
```

Var1	Freq
Feminino	0.5711462
Masculino	0.4288538

**1.5) [1 valor] Completação das frases seguintes:**

Inicialmente, o número de casos omissos na variável q1 era 5. No conjunto de dados em análise (depois de eliminar os registos com observações omissas) o número de estudantes trabalhadores é igual a **116**. A correlação mais elevada entre o pares de variáveis q14 tem o valor **0.600735876**. A correlação maior entre a variável q0 e as variáveis q19\_ regista-se para a variável q19\_\*\*3\*\*

2.1) [2 valores] Aprendizagem sobre  $CJ.train[,c(6:10)]$  e considerando  $y=CJ.train\$q19\_8$  recorrendo a one-hold-out validation; determinação de um “melhor” valor de K atendendo ao Sum of Squares Error

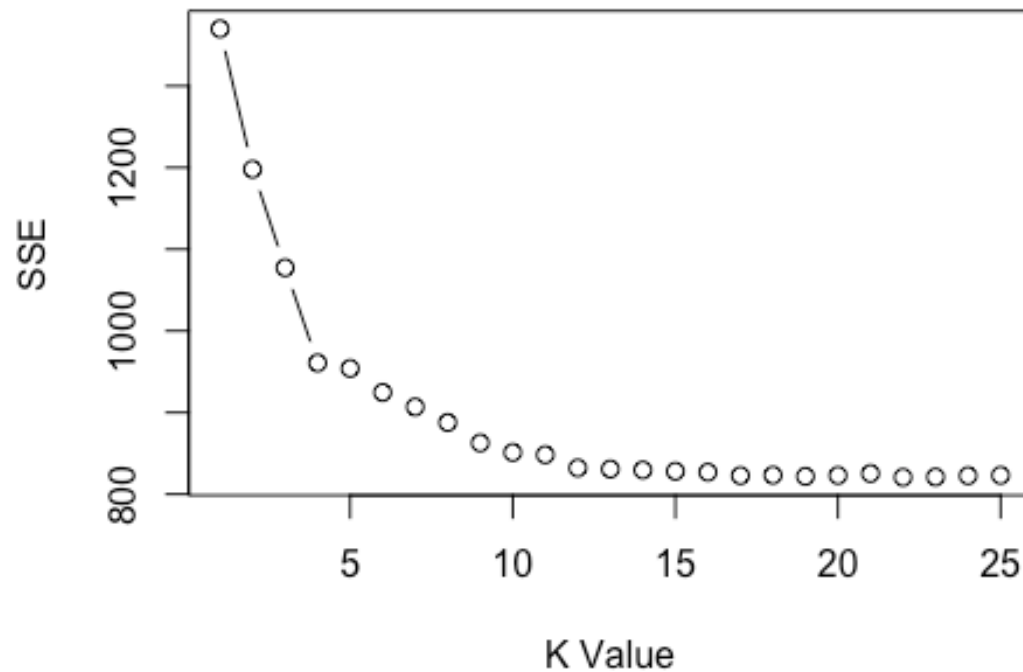
[illegible]

				q	q	q																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
--	--	--	--	---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

u	lin																																																																																			
g	o																																																																																			
al																																																																																				
1	P	Fe	1	Estudante	S	1	0	0	1	1	2	2	4	4	4	4	4	2	4	2	3	3	4	1	5	5	4	1																																																								
5	o	mi	9	a tempo	i																																																																															
rt	ni			inteiro	m																																																																															
u	no																																																																																			
g																																																																																				
al																																																																																				
1	P	Fe	2	Estudante	N	1	0	0	0	1	4	5	2	3	1	1	3	3	1	1	2	5	5	4	1	3	4	1																																																								
6	o	mi	0	a tempo	a																																																																															
rt	ni			inteiro	o																																																																															
u	no																																																																																			
g																																																																																				
al																																																																																				

```
#Selecionar o melhor k
k.sse<-matrix(NA,25,2)
for (i in 1:25){
  knn.CJ_k <- knn.reg(CJ.train_n[,c(6:10)], y=CJ.train_n$q19_8, k=i)
  k.sse[i,1]<-i
  k.sse[i,2] <- sse(CJ.train_n$q19_8, knn.CJ_k$pred)
}

#R-square plot
plot(k.sse[,2], type="b", xlab="K Value",ylab="SSE")
```



```
(k.sse_sort <- k.sse[order(k.sse[,2],decreasing = FALSE), ])
```

```
##      [,1]      [,2]
## [1,] 22 820.7169
## [2,] 23 821.0851
## [3,] 19 821.8892
## [4,] 24 822.6632
## [5,] 17 822.7474
## [6,] 20 823.0125
## [7,] 25 823.1408
## [8,] 18 823.3302
## [9,] 21 825.3696
## [10,] 16 826.9844
## [11,] 15 828.1111
## [12,] 14 829.6582
## [13,] 13 830.8107
## [14,] 12 832.1389
## [15,] 11 848.3471
## [16,] 10 851.1600
## [17,] 9 862.7901
## [18,] 8 887.8438
## [19,] 7 906.8980
## [20,] 6 924.6389
```

```
## [21,]    5  954.0400
## [22,]    4  960.8125
## [23,]    3 1077.3333
## [24,]    2 1198.0000
## [25,]    1 1370.0000

(best_k <- k.sse_sort[1,1])

## [1] 22
```

## 2.2) [2 valores] Considerando o “melhor” valor de K (v. 2.1), obtenção de estimativas do alvo e listagem dos 6 primeiros valores estimados nos conjuntos CJ.train e CJ.test

```
# estimativas sobre CJ.test
knn.CJ1 <- knn.reg(CJ.train_n[,c(6:10)], CJ.test_n[,c(6:10)],
y=CJ.train_n$q19_8, k=best_k)
knn.CJ1

## Prediction:
## [1] 3.227273 3.045455 3.227273 3.409091 3.090909 3.136364 3.454545
3.090909
## [9] 3.045455 3.045455 3.454545 3.227273 3.409091 2.818182 3.454545
3.681818
## [17] 3.409091 3.045455 3.090909 3.272727 3.681818 3.090909 2.772727
3.045455
## [25] 3.090909 2.818182 3.090909 3.136364 3.136364 3.227273 3.227273
3.454545
## [33] 3.227273 3.181818 3.045455 3.090909 3.409091 3.136364 3.227273
3.181818
## [41] 3.045455 3.681818 3.045455 3.318182 3.181818 3.090909 3.227273
2.818182
## [49] 3.136364 3.181818 3.227273 3.409091 3.090909 3.681818 3.681818
3.136364
## [57] 3.409091 3.090909 3.227273 3.090909 3.272727 3.681818 3.272727
3.318182
## [65] 3.272727 3.045455 3.136364 3.136364 3.136364 3.409091 3.272727
3.045455
## [73] 3.227273 3.090909 3.090909 3.318182 3.272727 3.409091 3.045455
3.681818
## [81] 3.090909 3.454545 3.090909 3.409091 3.454545 3.409091 3.681818
3.318182
## [89] 3.272727 3.227273 3.045455 3.136364 3.272727 3.409091 3.136364
3.090909
## [97] 3.045455 3.181818 3.136364 3.181818 3.681818 3.272727 3.181818
3.272727
## [105] 3.681818 3.227273 3.227273 3.136364 3.227273 3.090909 3.090909
3.136364
## [113] 3.454545 3.318182 3.454545 3.227273 3.090909 3.454545 3.136364
3.409091
## [121] 3.090909 3.227273 3.409091 3.136364 3.272727 3.136364 3.045455
```



3.181818  
## [129] 3.227273 3.409091 3.227273 3.227273 3.045455 2.818182 3.272727  
3.272727  
## [137] 3.045455 3.681818 3.045455 3.318182 3.045455 3.409091 3.409091  
3.181818  
## [145] 3.181818 3.181818 3.045455 3.181818 3.227273 3.136364 3.136364  
3.045455  
## [153] 3.409091 3.090909 3.454545 3.045455 3.045455 3.136364 3.090909  
3.227273  
## [161] 2.772727 3.136364 3.409091 3.090909 3.181818 3.181818 3.181818  
3.227273  
## [169] 2.772727 3.090909 3.136364 3.318182 3.318182 3.045455 3.045455  
3.136364  
## [177] 3.136364 3.409091 3.136364 2.772727 3.136364 3.045455 3.454545  
3.045455  
## [185] 3.181818 3.090909 3.136364 3.181818 3.136364 3.045455 2.772727  
3.090909  
## [193] 3.409091 3.090909 3.409091 3.090909 3.227273 3.454545 3.318182  
3.181818  
## [201] 3.090909 3.090909 3.090909 3.136364 3.136364 3.136364 3.681818  
3.681818  
## [209] 3.454545 3.318182 3.181818 3.090909 2.772727 3.136364 3.090909  
3.045455  
## [217] 3.136364 3.272727 3.090909 3.181818 3.045455 2.727273 3.045455  
3.090909  
## [225] 3.136364 3.136364 3.454545 3.045455 3.090909 3.136364 3.136364  
3.136364  
## [233] 3.136364 3.136364 3.181818 3.090909 2.772727 3.272727 3.681818  
3.227273  
## [241] 3.272727 3.045455 3.045455 3.136364 3.045455 3.318182 2.818182  
3.181818  
## [249] 3.181818 3.090909 3.090909 3.181818 3.181818 3.181818 2.818182  
3.318182  
## [257] 3.090909 3.136364 3.090909 3.000000 3.272727 3.181818 3.272727  
3.136364  
## [265] 3.272727 3.227273 3.409091 3.272727 3.409091 3.454545 3.181818  
2.772727  
## [273] 3.045455 3.045455 3.409091 3.090909 3.409091 3.409091 3.227273  
3.045455  
## [281] 2.772727 3.181818 3.045455 3.090909 3.454545 3.409091 3.045455  
3.454545  
## [289] 3.090909 3.136364 3.045455 3.136364 3.181818 3.272727 3.045455  
3.136364  
## [297] 3.045455 3.045455 3.045455 3.090909 3.045455 3.181818 3.181818  
3.045455  
## [305] 3.181818 3.272727 3.409091 3.181818 3.136364 3.045455 3.090909  
3.272727  
## [313] 3.045455 3.409091 3.045455 3.045455 3.227273 3.181818 3.136364  
2.818182  
## [321] 3.181818 2.772727 3.454545 3.090909 3.045455 3.409091 3.090909

```

3.409091
## [329] 3.181818 3.181818 3.045455 3.272727 3.136364 3.136364 3.409091
3.045455
## [337] 3.409091 3.045455 3.272727 3.454545 2.818182 3.272727 3.090909
3.409091
## [345] 3.045455 3.136364 3.681818 3.272727 2.818182 3.136364 3.045455
3.227273
## [353] 2.772727 3.272727 3.318182 3.090909 3.045455 3.409091 2.818182
3.090909
## [361] 2.772727 3.000000 3.409091 3.227273 3.181818 3.409091 3.272727
3.454545
## [369] 3.409091 3.181818 3.181818 3.045455 3.681818 3.181818 3.409091
3.181818
## [377] 3.181818 2.818182 3.045455 3.181818 3.318182 3.454545 3.318182
3.681818
## [385] 3.045455 3.272727 3.409091 3.409091 2.818182 3.181818 3.000000
3.681818
## [393] 2.727273 3.272727 3.272727 3.272727 3.136364 3.409091 3.045455
3.000000
## [401] 3.181818 2.772727 3.090909 3.272727 3.272727 2.772727 3.227273
3.181818
## [409] 3.045455 2.772727 3.181818 3.045455 2.772727 3.045455 3.045455
3.136364
## [417] 3.090909 3.136364 2.727273 2.772727 2.772727 2.727273 3.181818
3.090909
## [425] 3.090909 3.181818 2.772727 3.090909 3.090909 2.772727 2.772727
2.772727
## [433] 2.772727 2.772727 3.409091 3.181818 3.136364 3.181818 3.272727
3.045455
## [441] 2.818182 3.181818 3.090909 2.818182 3.409091 3.181818 3.045455
3.409091
## [449] 3.272727 3.409091 3.409091 2.772727 3.045455 3.090909 3.409091
3.090909
## [457] 3.181818 3.045455 3.409091 3.409091 3.409091 3.409091 3.272727
2.772727
## [465] 3.409091 3.409091 3.409091 3.000000 3.409091 3.409091 3.409091
3.409091
## [473] 3.409091 3.409091 3.045455 2.818182 3.409091 2.818182 3.409091
3.409091
## [481] 3.409091 3.409091 2.772727 3.227273 2.772727 3.409091 2.818182
3.045455
## [489] 3.409091 3.045455 3.454545 2.772727 3.454545 3.409091 3.409091
3.409091
## [497] 3.409091 3.409091 3.409091 3.409091 3.409091 3.409091 3.409091
3.409091
## [505] 3.136364 3.409091

```

```

# estimativas sobre CJ train

```

```

knn.CJ <- knn.reg(CJ.train_n[,c(6:10)], y=CJ.train_n$q19_8, k=best_k)
knn.CJ

```

```
knitr::kable(head(CJ.test_n, 6))
```

				q	q	q													q	q	q	q	q	q	q	q	
				1	1	1													1	1	1	1	1	1	1	1	
				2	2	2	q	q	q	q	q	q	q	q	q	q	q	1	1	1	1	1	1	1	1	1	
				q	b	b	b	1	1	1	1	1	1	1	1	1	1	9	9	9	9	9	9	9	9	9	
q				1	-	-	-	3	3	4	4	4	4	4	4	4	4	-	-	-	-	-	-	-	-	-	
0	q1	2	q3	0	a	b	c	a	b	a	b	c	d	e	f	g	h	i	j	1	2	3	4	5	6	7	8

g  
al

```
knitr::kable(head(CJ.train_n, 6))
```

					q	q	q													q	q	q	q	q	q	q	q	q	q	q	q	q
					1	1	1													1	1	1	1	1	1	1	1	1	1	1	1	1
					2	2	2	q	q	q	q	q	q	q	q	q	q	q	q	1	1	1	1	1	1	1	1					
				q	b	b	b	1	1	1	1	1	1	1	1	1	1	1	1	9	9	9	9	9	9	9	9					
q0	q	q	q3		1	-	-	-	3	3	4	4	4	4	4	4	4	4	-	-	-	-	-	-	-	-						
				0	a	b	c	a	b	a	b	c	d	e	f	g	h	i	j	1	2	3	4	5	6	7	8					
9	M	Fe	2	N	1	1	1	0	0	5	4	4	5	3	3	3	2	2	2	3	4	3	4	1	4	3	4					
1	ac	m	3	a																												
5	au	in		o																												
		in																														
		o																														
1	Ch	M	2	S	1	1	0	1	1	3	2	2	4	5	3	2	4	4	4	4	4	4	1	4	2	3						
1	in	as	1	i																												
3	a	cu		m																												
2		li																														
		no																														
1	M	M	2	S	0	1	0	1	1	3	4	2	5	3	1	3	1	4	1	2	3	1	4	4	3	4						
3	oc	as	1	i																												
8	a	cu		m																												
4	m	li																														
		no																														
		qu																														
		e																														
1	Ch	M	2	N	1	0	0	0	0	3	3	2	4	2	4	4	1	1	1	4	2	2	5	1	2	1	4					
1	in	as	1	a																												
0	a	cu		o																												
5		li																														
		no																														
4	Po	Fe	2	N	0	0	1	0	0	4	5	4	3	3	3	3	3	2	2	4	4	1	4	2	4	2	5					
2	rt	m	4	a																												
4	ug	in		o																												
	al	in																														
		o																														
1	Po	Fe	1	S	1	1	1	0	0	3	4	3	3	3	3	3	2	3	2	3	3	4	2	2	3	3	5					

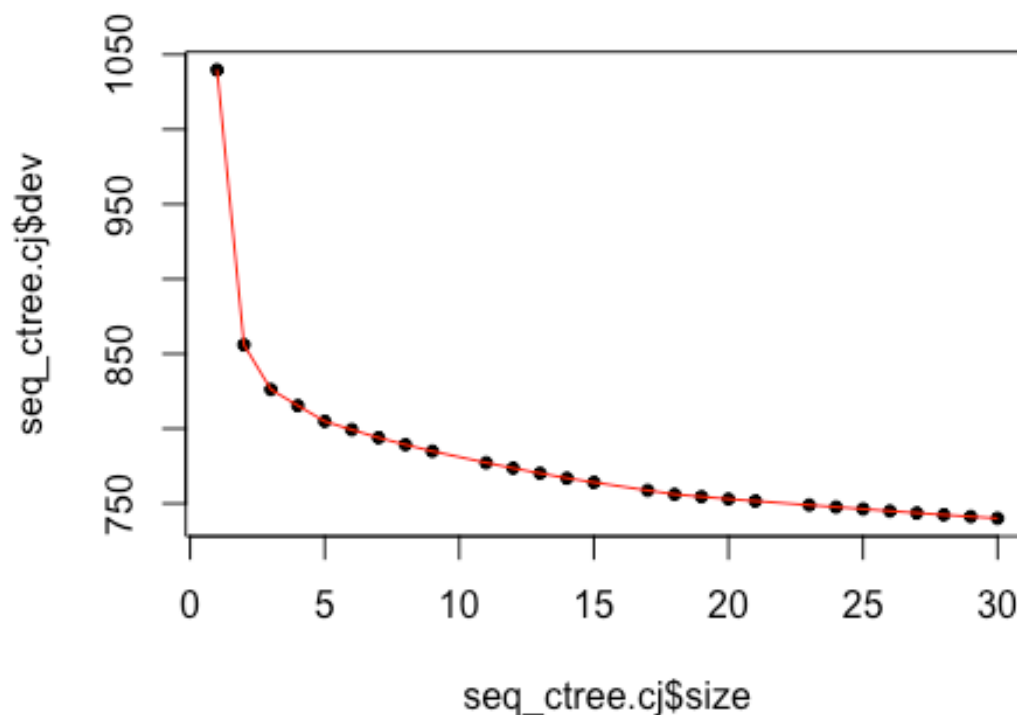
	q0	q1	q2	q3		q1	q2	q3		q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12	q13	q14	q15	q16	q17	q18	q19	q20	q21	q22	q23	q24	q25	q26	q27	q28	q29	q30	q31	q32	q33	q34	q35	q36	q37	q38	q39	q40	q41	q42	q43	q44	q45	q46	q47	q48	q49	q50	q51	q52	q53	q54	q55	q56	q57	q58	q59	q60	q61	q62	q63	q64	q65	q66	q67	q68	q69	q70	q71	q72	q73	q74	q75	q76	q77	q78	q79	q80	q81	q82	q83	q84	q85	q86	q87	q88	q89	q90	q91	q92	q93	q94	q95	q96	q97	q98	q99	q100	q101	q102	q103	q104	q105	q106	q107	q108	q109	q110	q111	q112	q113	q114	q115	q116	q117	q118	q119	q120	q121	q122	q123	q124	q125	q126	q127	q128	q129	q130	q131	q132	q133	q134	q135	q136	q137	q138	q139	q140	q141	q142	q143	q144	q145	q146	q147	q148	q149	q150	q151	q152	q153	q154	q155	q156	q157	q158	q159	q160	q161	q162	q163	q164	q165	q166	q167	q168	q169	q170	q171	q172	q173	q174	q175	q176	q177	q178	q179	q180	q181	q182	q183	q184	q185	q186	q187	q188	q189	q190	q191	q192	q193	q194	q195	q196	q197	q198	q199	q200	q201	q202	q203	q204	q205	q206	q207	q208	q209	q210	q211	q212	q213	q214	q215	q216	q217	q218	q219	q220	q221	q222	q223	q224	q225	q226	q227	q228	q229	q230	q231	q232	q233	q234	q235	q236	q237	q238	q239	q240	q241	q242	q243	q244	q245	q246	q247	q248	q249	q250	q251	q252	q253	q254	q255	q256	q257	q258	q259	q260	q261	q262	q263	q264	q265	q266	q267	q268	q269	q270	q271	q272	q273	q274	q275	q276	q277	q278	q279	q280	q281	q282	q283	q284	q285	q286	q287	q288	q289	q290	q291	q292	q293	q294	q295	q296	q297	q298	q299	q300	q301	q302	q303	q304	q305	q306	q307	q308	q309	q310	q311	q312	q313	q314	q315	q316	q317	q318	q319	q320	q321	q322	q323	q324	q325	q326	q327	q328	q329	q330	q331	q332	q333	q334	q335	q336	q337	q338	q339	q340	q341	q342	q343	q344	q345	q346	q347	q348	q349	q350	q351	q352	q353	q354	q355	q356	q357	q358	q359	q360	q361	q362	q363	q364	q365	q366	q367	q368	q369	q370	q371	q372	q373	q374	q375	q376	q377	q378	q379	q380	q381	q382	q383	q384	q385	q386	q387	q388	q389	q390	q391	q392	q393	q394	q395	q396	q397	q398	q399	q400	q401	q402	q403	q404	q405	q406	q407	q408	q409	q410	q411	q412	q413	q414	q415	q416	q417	q418	q419	q420	q421	q422	q423	q424	q425	q426	q427	q428	q429	q430	q431	q432	q433	q434	q435	q436	q437	q438	q439	q440	q441	q442	q443	q444	q445	q446	q447	q448	q449	q450	q451	q452	q453	q454	q455	q456	q457	q458	q459	q460	q461	q462	q463	q464	q465	q466	q467	q468	q469	q470	q471	q472	q473	q474	q475	q476	q477	q478	q479	q480	q481	q482	q483	q484	q485	q486	q487	q488	q489	q490	q491	q492	q493	q494	q495	q496	q497	q498	q499	q500	q501	q502	q503	q504	q505	q506	q507	q508	q509	q510	q511	q512	q513	q514	q515	q516	q517	q518	q519	q520	q521	q522	q523	q524	q525	q526	q527	q528	q529	q530	q531	q532	q533	q534	q535	q536	q537	q538	q539	q540	q541	q542	q543	q544	q545	q546	q547	q548	q549	q550	q551	q552	q553	q554	q555	q556	q557	q558	q559	q560	q561	q562	q563	q564	q565	q566	q567	q568	q569	q570	q571	q572	q573	q574	q575	q576	q577	q578	q579	q580	q581	q582	q583	q584	q585	q586	q587	q588	q589	q590	q591	q592	q593	q594	q595	q596	q597	q598	q599	q600	q601	q602	q603	q604	q605	q606	q607	q608	q609	q610	q611	q612	q613	q614	q615	q616	q617	q618	q619	q620	q621	q622	q623	q624	q625	q626	q627	q628	q629	q630	q631	q632	q633	q634	q635	q636	q637	q638	q639	q640	q641	q642	q643	q644	q645	q646	q647	q648	q649	q650	q651	q652	q653	q654	q655	q656	q657	q658	q659	q660	q661	q662	q663	q664	q665	q666	q667	q668	q669	q670	q671	q672	q673	q674	q675	q676	q677	q678	q679	q680	q681	q682	q683	q684	q685	q686	q687	q688	q689	q690	q691	q692	q693	q694	q695	q696	q697	q698	q699	q700	q701	q702	q703	q704	q705	q706	q707	q708	q709	q710	q711	q712	q713	q714	q715	q716	q717	q718	q719	q720	q721	q722	q723	q724	q725	q726	q727	q728	q729	q730	q731	q732	q733	q734	q735	q736	q737	q738	q739	q740	q741	q742	q743	q744	q745	q746	q747	q748	q749	q750	q751	q752	q753	q754	q755	q756	q757	q758	q759	q760	q761	q762	q763	q764	q765	q766	q767	q768	q769	q770	q771	q772	q773	q774	q775	q776	q777	q778	q779	q780	q781	q782	q783	q784	q785	q786	q787	q788	q789	q790	q791	q792	q793	q794	q795	q796	q797	q798	q799	q800	q801	q802	q803	q804	q805	q806	q807	q808	q809	q810	q811	q812	q813	q814	q815	q816	q817	q818	q819	q820	q821	q822	q823	q824	q825	q826	q827	q828	q829	q830	q831	q832	q833	q834	q835	q836	q837	q838	q839	q840	q841	q842	q843	q844	q845	q846	q847	q848	q849	q850	q851	q852	q853	q854	q855	q856	q857	q858	q859	q860	q861	q862	q863	q864	q865	q866	q867	q868	q869	q870	q871	q872	q873	q874	q875	q876	q877	q878	q879	q880	q881	q882	q883	q884	q885	q886	q887	q888	q889	q890	q891	q892	q893	q894	q895	q896	q897	q898	q899	q900	q901	q902	q903	q904	q905	q906	q907	q908	q909	q910	q911	q912	q913	q914	q915	q916	q917	q918	q919	q920	q921	q922	q923	q924	q925	q926	q927	q928	q929	q930	q931	q932	q933	q934	q935	q936	q937	q938	q939	q940	q941	q942	q943	q944	q945	q946	q947	q948	q949	q950	q951	q952	q953	q954	q955	q956	q957	q958	q959	q960	q961	q962	q963	q964	q965	q966	q967	q968	q969	q970	q971	q972	q973	q974	q975	q976	q977	q978	q979	q980	q981	q982	q983	q984	q985	q986	q987	q988	q989	q990	q991	q992	q993	q994	q995	q996	q997	q998	q999	q1000	q1001	q1002	q1003	q1004	q1005	q1006	q1007	q1008	q1009	q1010	q1011	q1012	q1013	q1014	q1015	q1016	q1017	q1018	q1019	q1020	q1021	q1022	q1023	q1024	q1025	q1026	q1027	q1028	q1029	q1030	q1031	q1032	q1033	q1034	q1035	q1036	q1037	q1038	q1039	q1040	q1041	q1042	q1043	q1044	q1045	q1046	q1047	q1048	q1049	q1050	q1051	q1052	q1053	q1054	q1055	q1056	q1057	q1058	q1059	q1060	q1061	q1062	q1063	q1064	q1065	q1066	q1067	q1068	q1069	q1070	q1071	q1072	q1073	q1074	q1075	q1076	q1077	q1078	q1079	q1080	q1081	q1082	q1083	q1084	q1085	q1086	q1087	q1088	q1089	q1090	q1091	q1092	q1093	q1094	q1095	q1096	q1097	q1098	q1099	q1100	q1101	q1102	q1103	q1104	q1105	q1106	q1107	q1108	q1109	q1110	q1111	q1112	q1113	q1114	q1115	q1116	q1117	q1118	q1119	q1120	q1121	q1122	q1123	q1124	q1125	q1126	q1127	q1128	q1129	q1130	q1131	q1132	q1133	q1134	q1135	q1136	q1137	q1138	q1139	q1140	q1141	q1142	q1143	q1144	q1145	q1146	q1147	q1148	q1149	q1150	q1151	q1152	q1153	q1154	q1155	q1156	q1157	q1158	q1159	q1160	q1161	q1162	q1163	q1164	q1165	q1166	q1167	q1168	q1169	q1170	q1171	q1172	q1173	q1174	q1175	q1176	q1177	q1178	q1179	q1180	q1181	q1182	q1183	q1184	q1185	q1186	q1187	q1188	q1189	q1190	q1191	q1192	q1193	q1194	q1195	q1196	q1197	q1198	q1199	q1200	q1201	q1202	q1203	q1204	q1205	q1206	q1207	q1208	q1209	q1210	q1211	q1212	q1213	q1214	q1215	q1216	q1217	q1218	q1219	q1220	q1221	q1222	q1223	q1224	q1225	q1226	q1227	q1228	q1229	q1230	q1231	q1232	q1233	q1234	q1235	q1236	q1237	q1238	q1239	q1240	q1241	q1242	q1243	q1244	q1245	q1246	q1247	q1248	q1249	q1250	q1251	q1252	q1253	q1254	q1255	q1256	q1257	q1258	q1259	q1260	q1261	q1262	q1263	q1264	q1265	q1266	q1267	q1268	q1269	q1270	q1271	q1272	q1273	q1274	q1275	q1276	q1277	q1278	q1279	q1280	q1281	q1282	q1283	q1284	q1285	q1286	q1287	q1288	q1289	q1290	q1291	q1292	q1293	q1294	q1295	q1296	q1297	q1298	q1299	q1300	q1301	q1302	q1303	q1304	q1305	q1306	q1307	q1308	q1309	q1310	q1311	q1312	q1313	q1314	q1315	q1316	q1317	q1318	q1319	q1320	q1321	q1322	q1323	q1324	q1325	q1326	q1327	q1328	q1329	q1330	q1331	q1332	q1333	q1334	q1335	q1336	q1337	q1338	q1339	q13
--	----	----	----	----	--	----	----	----	--	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-----

### 3. Classificação: utilização de uma Árvore para prever q10 (Compra ou não compra produtos de marca) considerando 4 preditores: q12b\_a, q13a, q14e e q14i.

#### 3.1) [2 valores] Construção de uma Árvore de classificação sobre CJ.train efetuando a sua poda de modo a fixar 15 nós folha (para prever q10 com base nos preditores q12b\_a, q13a, q14e e q14i)

```
ctree_large.cj<-tree(q10~ q12b_a+ q13a+ q14e + q14i,data=CJ.train,  
control=tree.control(nrow(CJ.train),  
mincut = 1, minsize = 2, mindev = 0.001), split = "deviance")
```

```
# cost-complexity pruning  
seq_ctree.cj <- prune.tree(ctree_large.cj)  
plot(seq_ctree.cj$size, seq_ctree.cj$dev, pch=20)  
lines(seq_ctree.cj$size, seq_ctree.cj$dev, col="red")
```



```
ctree_large.cj<-prune.tree(ctree_large.cj, best=15 )  
summary(ctree_large.cj)
```

```
##
## Classification tree:
## snip.tree(tree = ctree_large.cj, nodes = c(74L, 75L, 12L, 44L,
## 8L, 46L, 47L, 10L, 14L))
## Number of terminal nodes: 15
## Residual mean deviance: 1.027 = 764.1 / 744
## Misclassification error rate: 0.2437 = 185 / 759
```

### 3.2) [2 valores] Representações da Árvore de Classificação: a) Lista indentada; b) Gráfico da Árvore

# a)

```
print(ctree_large.cj)

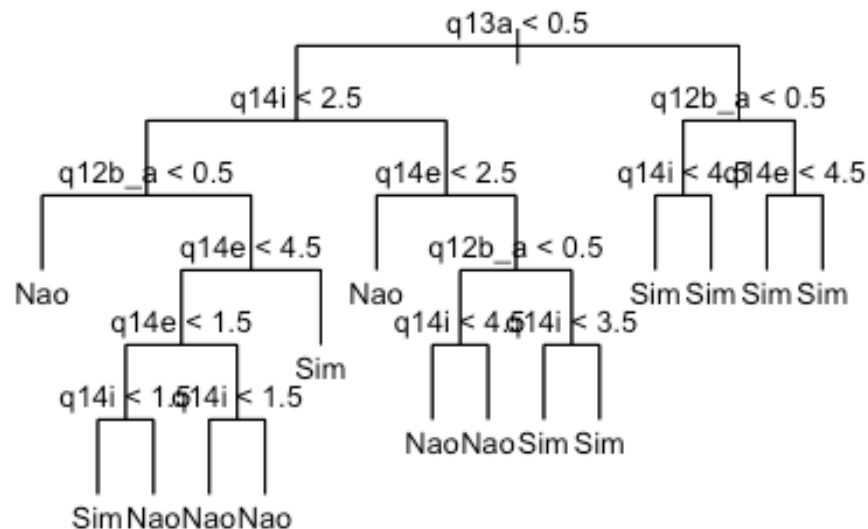
## node), split, n, deviance, yval, (yprob)
##      * denotes terminal node
##
## 1) root 759 1040.000 Sim ( 0.43610 0.56390 )
##    2) q13a < 0.5 450 592.500 Nao ( 0.63111 0.36889 )
##      4) q14i < 2.5 246 279.900 Nao ( 0.74390 0.25610 )
##        8) q12b_a < 0.5 133 122.500 Nao ( 0.82707 0.17293 ) *
##        9) q12b_a > 0.5 113 146.900 Nao ( 0.64602 0.35398 )
##        18) q14e < 4.5 108 137.500 Nao ( 0.66667 0.33333 )
##          36) q14e < 1.5 10 13.860 Nao ( 0.50000 0.50000 )
##            72) q14i < 1.5 8 10.590 Sim ( 0.37500 0.62500 ) *
##            73) q14i > 1.5 2 0.000 Nao ( 1.00000 0.00000 ) *
##          37) q14e > 1.5 98 122.300 Nao ( 0.68367 0.31633 )
##            74) q14i < 1.5 20 13.000 Nao ( 0.90000 0.10000 ) *
##            75) q14i > 1.5 78 102.900 Nao ( 0.62821 0.37179 ) *
##        19) q14e > 4.5 5 5.004 Sim ( 0.20000 0.80000 ) *
##    5) q14i > 2.5 204 282.800 Sim ( 0.49510 0.50490 )
##      10) q14e < 2.5 41 52.640 Nao ( 0.65854 0.34146 ) *
##      11) q14e > 2.5 163 224.600 Sim ( 0.45399 0.54601 )
##        22) q12b_a < 0.5 72 98.920 Nao ( 0.55556 0.44444 )
##        44) q14i < 4.5 69 95.290 Nao ( 0.53623 0.46377 ) *
##        45) q14i > 4.5 3 0.000 Nao ( 1.00000 0.00000 ) *
##        23) q12b_a > 0.5 91 120.300 Sim ( 0.37363 0.62637 )
##        46) q14i < 3.5 68 92.790 Sim ( 0.42647 0.57353 ) *
##        47) q14i > 3.5 23 24.080 Sim ( 0.21739 0.78261 ) *
##    3) q13a > 0.5 309 263.500 Sim ( 0.15210 0.84790 )
##      6) q12b_a < 0.5 117 128.800 Sim ( 0.23932 0.76068 )
##        12) q14i < 4.5 112 126.000 Sim ( 0.25000 0.75000 ) *
##        13) q14i > 4.5 5 0.000 Sim ( 0.00000 1.00000 ) *
##    7) q12b_a > 0.5 192 124.000 Sim ( 0.09896 0.90104 )
##      14) q14e < 4.5 171 119.300 Sim ( 0.11111 0.88889 ) *
##      15) q14e > 4.5 21 0.000 Sim ( 0.00000 1.00000 ) *
```

# b)

```
plot(ctree_large.cj, type="uniform")
```

```
text(ctree_large.cj,pretty =0,cex=0.8)
title(main = "Prunned Classification Tree for q10")
```

## Prunned Classification Tree for q10



### 3.3) [2 valores] Obtenção, sobre as amostras CJ.train e CJ.test, das “Matrizes de Confusão” e correspondentes medidas Accuracy associadas à Árvore de Classificação

```
# "Matriz de confusão" e accuracy sobre CJ.train
pred.ctree <- predict(ctree_large.cj, CJ.train,type="class")
(confusion_mat <- table(CJ.train$q10, pred.ctree))

##      pred.ctree
##      Nao Sim
##  Nao 246  85
##  Sim 100 328

#Accuracy
(accuracy<-sum(diag(confusion_mat))/sum(confusion_mat))

## [1] 0.7562582

# "Matriz de confusão" e accuracy sobre CJ.test
```



```

pred.ctree1 <- predict(ctree_large.cj, CJ.test, type="class")
(confusion_mat1 <- table(CJ.test$q10, pred.ctree1))

##      pred.ctree1
##      Nao Sim
## Nao 158  67
## Sim  78 203

#Accuracy
(accuracy1<-sum(diag(confusion_mat1))/sum(confusion_mat1))

## [1] 0.7134387

```

### 3.4) [1 valor] Completação das frases seguintes:

A árvore obtida, classifica as observações do nó folha 73) na classe **\*\*\_Não; o nó folha com o maior número de observações de treino é o nó \_\_14; no conjunto de teste o número de observações corretamente classificadas nas classes “Não” e “Sim” é \_158\*\* e 203\_.** respectivamente.