

Prevendo Despesas Hospitalares_v3

Evanil Tiengo Junior

Índice

Descrição:.....	1
Objetivo:.....	1
Algoritmo:	2
Etapa 1 - Coletando os Dados	3
Etapa 2 - EDA (Exploratory Data Analysis).....	3
Etapa 3 - Modelagem	28
Etapa 4 - Interpretando o Modelo	57
Etapa 5 - Previsão	60
Etapa 6 - Avaliando a Performance.....	62
Conclusão	63

Descrição:

Este experimento visa demonstrar o processo de construção de um modelo de regressão linear múltipla para prever os gastos de despesas hospitalares. Usaremos um conjunto de dados para construir e treinar nosso modelo

Este projeto faz parte da formação Big Data Analytics com R e Microsoft Azure da DSA. Para esta análise, foi usado um conjunto de dados simulando despesas médicas hipotéticas para um conjunto de pacientes espalhados por 4 regiões do Brasil. Esse dataset possui 1.338 observações e 7 variáveis. Todo o projeto será descrito de acordo com suas etapas.

Nesta versão 3, estaremos utilizando técnicas de normalização, correlação, tratamento de outliers para melhorar a performance do algoritmo.

Objetivo:

O objetivo será prever o valor da variável gasto com o menor erro possível.

Algoritmo:

```
# Local armazenamento
setwd("~/Mini-Projeto03")
getwd()

# Pacotes utilizados.
#install.packages("dplyr")
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("psych")
library(psych)

##
## Attaching package: 'psych'

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

#install.packages("corrplot")
library(corrplot)

## corrplot 0.84 loaded

#install.packages("caTools")
library(caTools)
#install.packages("car")
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

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

```
## The following object is masked from 'package:dplyr':
##
##      recode

#install.packages("caret")
library(caret)

## Loading required package: lattice
```

Etapa 1 - Coletando os Dados

Os dados foram fornecidos pela DSA. Sendo assim preciso realizar a carga. Antes de realizar a carga do arquivo é necessário saber o formato do mesmo, que neste caso é .csv!

```
Despesas <- read.csv("~/Mini-Projeto03/despesas.csv", stringsAsFactors = FALSE)
```

Etapa 2 - EDA (Exploratory Data Analysis)

Com o comando abaixo é possível identificar as classes de cada variável
str(Despesas)

```
## 'data.frame':    1338 obs. of  7 variables:
## $ idade   : int   19 18 28 33 32 31 46 37 37 60 ...
## $ sexo    : chr   "mulher" "homem" "homem" "homem" ...
## $ bmi      : num   27.9 33.8 33 22.7 28.9 25.7 33.4 27.7 29.8 25.8 ...
## $ filhos  : int    0 1 3 0 0 0 1 3 2 0 ...
## $ fumante : chr    "sim" "nao" "nao" "nao" ...
## $ regioao : chr    "sudeste" "sul" "sul" "nordeste" ...
## $ gastos  : num  16885 1726 4449 21984 3867 ...
```

Identificação de NA's e Vazios!

```
any(is.na(Despesas))
```

```
## [1] FALSE
```

```
any(Despesas == "")
```

```
## [1] FALSE
```

False significa que não existe nenhum campo com NA ou Vazio!

Para a previsão irei utilizar a Regressão Linear Múltipla. Sendo assim, irei transformar as variáveis em numéricas.

```
# $sexo: mulher = 1; homem = 2
```

```
Despesas$sexo <- as.factor(Despesas$sexo)
```

```

Despesas$sexo <- as.numeric(Despesas$sexo)

# $fumante: sim = 1; não = 2
Despesas$fumante <- as.factor(Despesas$fumante)
Despesas$fumante <- as.numeric(Despesas$fumante)

# $regiao: nordeste = 1; norte = 2; sudeste = 3; sul = 4
Despesas$regiao <- as.factor(Despesas$regiao)
Despesas$regiao <- as.numeric(Despesas$regiao)

# Abaixo temos um resumo dos dados
summary(Despesas)

##      idade      sexo      bmi      filhos
## Min.   :18.00  Min.   :1.000  Min.   :16.00  Min.   :0.000
## 1st Qu.:27.00  1st Qu.:1.000  1st Qu.:26.30  1st Qu.:0.000
## Median :39.00  Median :1.000  Median :30.40  Median :1.000
## Mean   :39.21  Mean   :1.495  Mean   :30.67  Mean   :1.095
## 3rd Qu.:51.00  3rd Qu.:2.000  3rd Qu.:34.70  3rd Qu.:2.000
## Max.   :64.00  Max.   :2.000  Max.   :53.10  Max.   :5.000
##      fumante      regiao      gastos
## Min.   :1.000  Min.   :1.000  Min.   : 1122
## 1st Qu.:1.000  1st Qu.:2.000  1st Qu.: 4740
## Median :1.000  Median :3.000  Median : 9382
## Mean   :1.205  Mean   :2.544  Mean   :13270
## 3rd Qu.:1.000  3rd Qu.:4.000  3rd Qu.:16640
## Max.   :2.000  Max.   :4.000  Max.   :63770

# Aqui podemos tirar algumas observações:

```

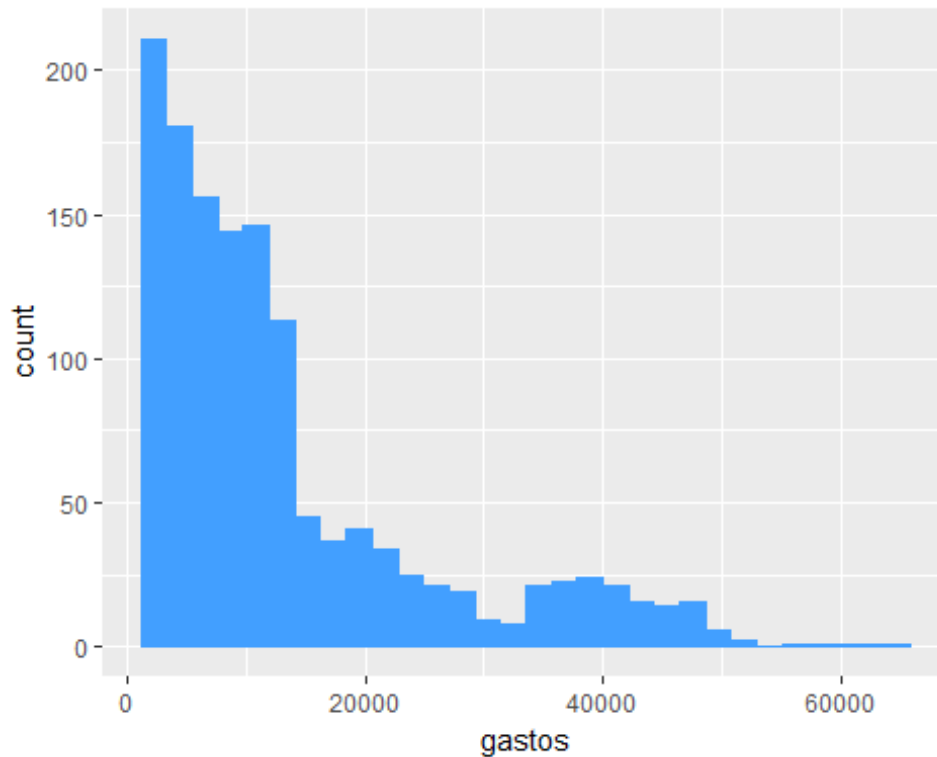
Agora vamos analisar a distribuição dos dados de forma gráfica:

```

# Medidas da tendência central da variável $gastos
ggplot(Despesas, aes(gastos)) +
  geom_histogram(color = "#429FFF", fill = "#429FFF")

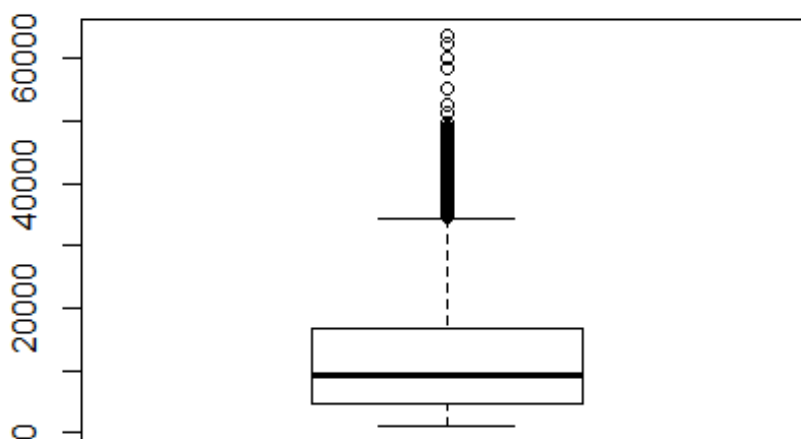
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```



Pelo histograma podemos identificar uma distribuição Unimodal e um enviesamento a direita (Mediana < Media). Com isso podemos observar que a distribuição dos dados concentra-se em torno dos valores mínimos e temos alguns outliers.

```
BoxPlot_Despesas <- boxplot(Despesas$gastos)
```



Como pode ser observado temos muitos outliers superiores.

```
# Estatísticas do boxplot  
BoxPlot_Despesas$stats
```

```
##      [,1]
## [1,] 1121.87
## [2,] 4738.27
## [3,] 9382.03
## [4,] 16657.72
## [5,] 34472.84

# 1 = Valor min
# 2 = 1 quartil
# 3 = mediana
# 4 = 3 quartil
# 5 = Valor max antes dos outliers

# Outliers
BoxPlot_Despesas$out

## [1] 39611.76 36837.47 37701.88 38711.00 35585.58 51194.56 39774.28
## [8] 48173.36 38709.18 37742.58 47496.49 37165.16 39836.52 43578.94
## [15] 47291.06 47055.53 39556.49 40720.55 36950.26 36149.48 48824.45
## [22] 43753.34 37133.90 34779.62 38511.63 35160.13 47305.31 44260.75
## [29] 41097.16 43921.18 36219.41 46151.12 42856.84 48549.18 47896.79
## [36] 42112.24 38746.36 42124.52 34838.87 35491.64 42760.50 47928.03
## [43] 48517.56 41919.10 36085.22 38126.25 42303.69 46889.26 46599.11
## [50] 39125.33 37079.37 35147.53 48885.14 36197.70 38245.59 48675.52
## [57] 63770.43 45863.21 39983.43 45702.02 58571.07 43943.88 39241.44
## [64] 42969.85 40182.25 34617.84 42983.46 42560.43 40003.33 45710.21
## [71] 46200.99 46130.53 40103.89 34806.47 40273.65 44400.41 40932.43
## [78] 40419.02 36189.10 44585.46 43254.42 36307.80 38792.69 55135.40
## [85] 43813.87 39597.41 36021.01 45008.96 37270.15 42111.66 40974.16
## [92] 46113.51 46255.11 44202.65 48673.56 35069.37 39047.29 47462.89
## [99] 38998.55 41999.52 41034.22 36580.28 35595.59 42211.14 44423.80
## [106] 37484.45 39725.52 44501.40 39727.61 48970.25 39871.70 34672.15
## [113] 41676.08 44641.20 41949.24 36124.57 38282.75 46661.44 40904.20
## [120] 36898.73 52590.83 40941.29 39722.75 37465.34 36910.61 38415.47
## [127] 41661.60 60021.40 47269.85 49577.66 37607.53 47403.88 38344.57
## [134] 34828.65 62592.87 46718.16 37829.72 36397.58 43896.38

# Valor mínimo do outliers
min(BoxPlot_Despesas$out)

## [1] 34617.84
```

Como pode ser observado, o 3rd Quartil que representa 75% dos valores, é igual a 16657.72 reais. Isso quer dizer que 75% dos gastos são inferiores a esse valor. Como os valores dos outliers são elevados não irei retirá-los do modelo nesse momento. Mais abaixo farei um teste retirando os outliers para saber a influência deles.

O valor máximo da variável gastos antes de se tornar outliers é 34617.84

Como a variável possui muitos outliers irei Normalizá-los pela regra do log.

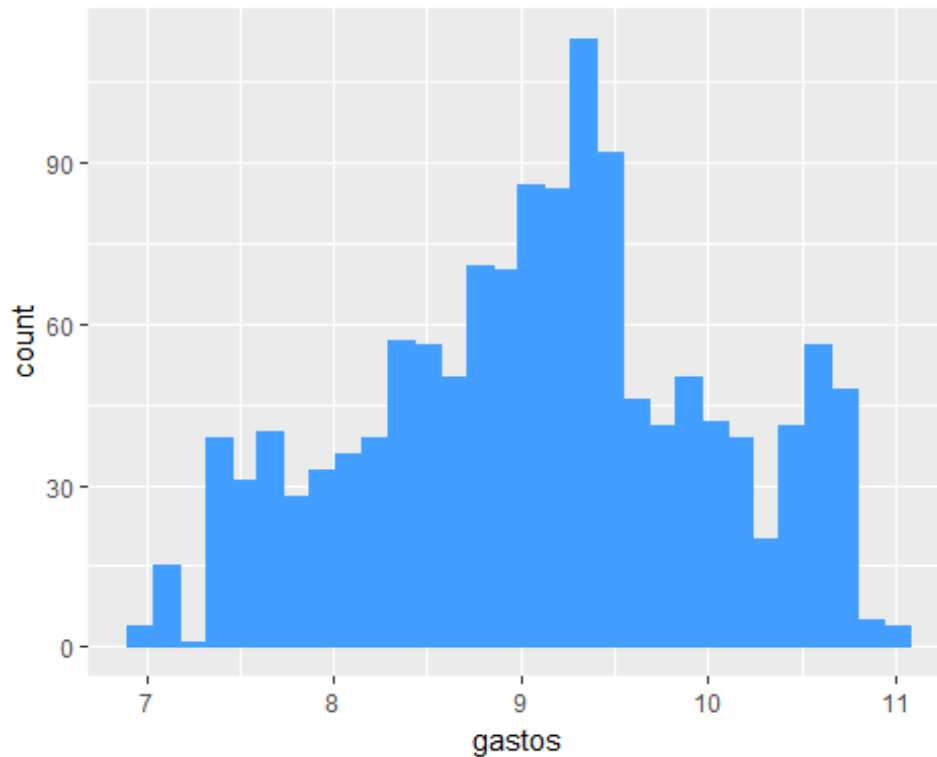
```
# Normalização da variável $gastos
Dados_Norm <- Despesas
Dados_Norm$gastos <- log(Dados_Norm$gastos)
str(Dados_Norm)

## 'data.frame': 1338 obs. of 7 variables:
## $ idade : int 19 18 28 33 32 31 46 37 37 60 ...
## $ sexo : num 2 1 1 1 1 2 2 2 1 2 ...
## $ bmi : num 27.9 33.8 33 22.7 28.9 25.7 33.4 27.7 29.8 25.8 ...
## $ filhos : int 0 1 3 0 0 0 1 3 2 0 ...
## $ fumante: num 2 1 1 1 1 1 1 1 1 1 ...
## $ regioao : num 3 4 4 1 1 4 4 1 2 1 ...
## $ gastos : num 9.73 7.45 8.4 10 8.26 ...
```

Vamos ver como ficou a variável \$gastos após a normalização.

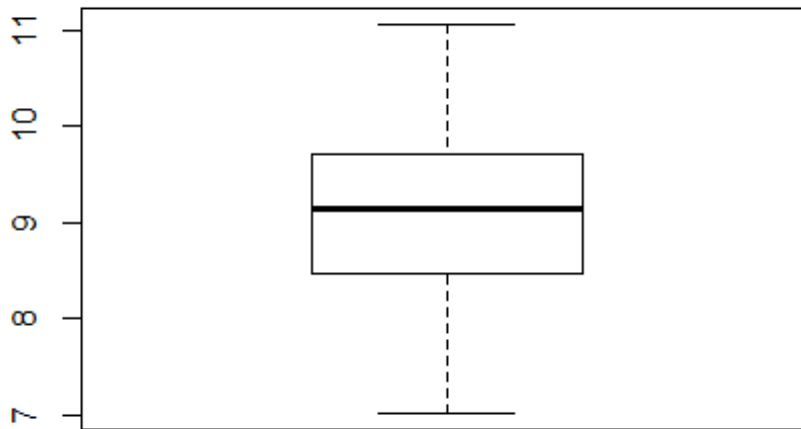
```
# Medidas da tendência central da variável $gastos
ggplot(Dados_Norm, aes(gastos)) +
  geom_histogram(color = "#429FFF", fill = "#429FFF")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Pelo histograma podemos identificar uma distribuição parecida com uma normal.

```
BoxPlot_Despesas <- boxplot(Dados_Norm$gastos)
```



Como pode ser observado não temos mais os outliers superiores.

```
# Estatísticas do boxplot
BoxPlot_Despesas$stats

##           [,1]
## [1,]  7.022752
## [2,]  8.463427
## [3,]  9.146551
## [4,]  9.720629
## [5,] 11.063045

# 1 = Valor min
# 2 = 1 quartil
# 3 = mediana
# 4 = 3 quartil
# 5 = Valor max antes dos outliers

# Outlier
BoxPlot_Despesas$out

## numeric(0)

# Valor mínimo do outliers
min(BoxPlot_Despesas$out)

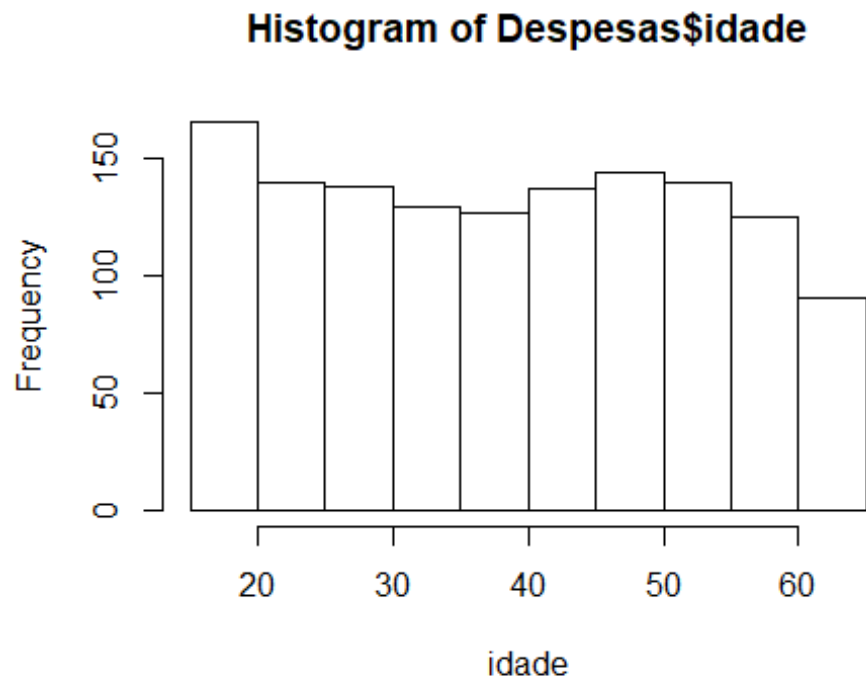
## Warning in min(BoxPlot_Despesas$out): nenhum argumento não faltante pa
ra
## min; retornando Inf

## [1] Inf
```

Agora os valores estão melhor distribuídos e isso torna o modelo ainda mais genérico.

Gráfico e Tabela de contingência para analisar a distribuição das outras variáveis.


```
# $idade: distribuição uniforme  
hist(Despesas$idade, xlab = "idade")
```

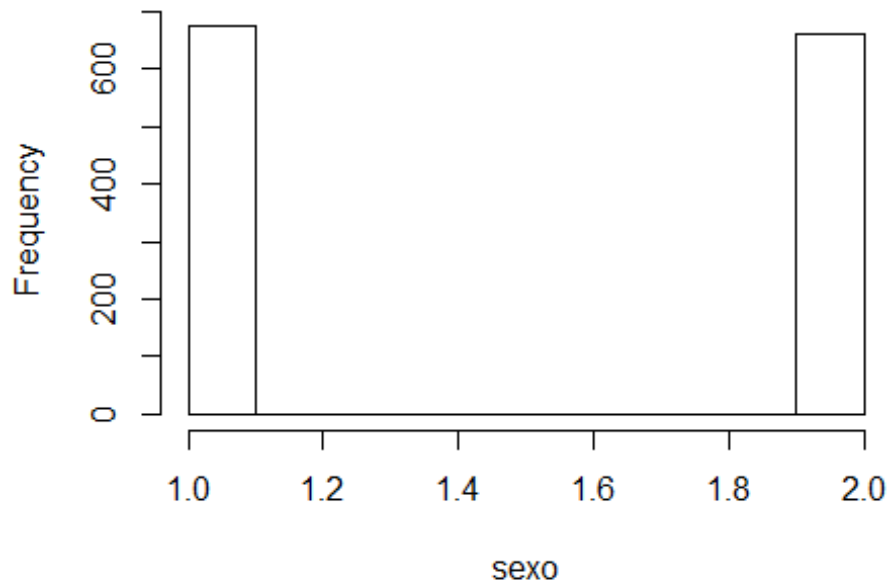


```
count(Despesas, idade)
```

```
## # A tibble: 47 x 2  
##   idade     n  
##   <int> <int>  
## 1     18    69  
## 2     19    68  
## 3     20    29  
## 4     21    28  
## 5     22    28  
## 6     23    28  
## 7     24    28  
## 8     25    28  
## 9     26    28  
## 10    27    28  
## # ... with 37 more rows
```

```
# $sexo: distribuição Dicotômica.  
hist(Despesas$sexo, xlab = "sexo")
```

Histogram of Despesas\$sexo



```
round(prop.table(table(Despesas$sexo)) * 100, digits = 1)
```

```
##
```

```
##      1      2
```

```
## 50.5 49.5
```

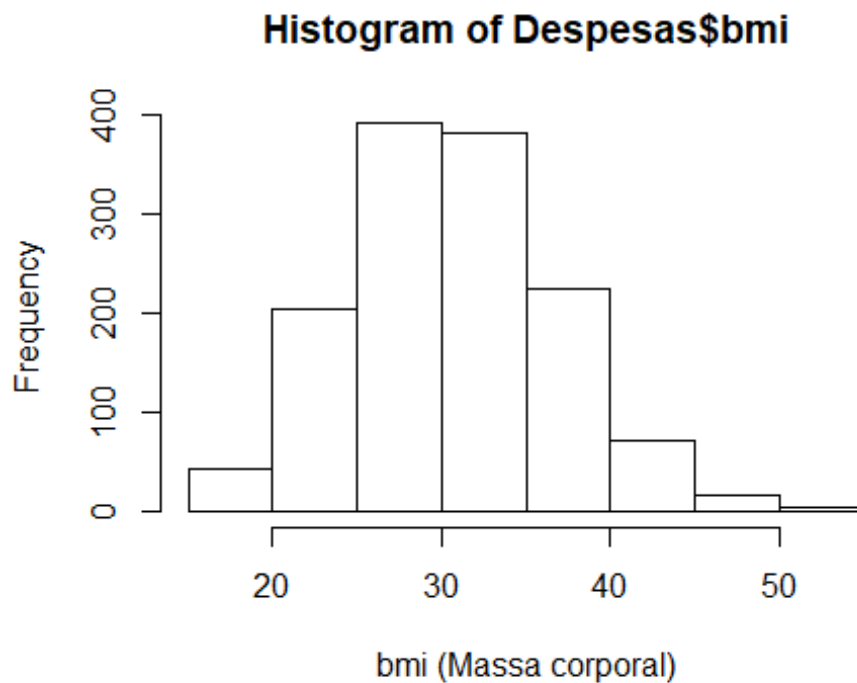
```
# Em termos de porcentagem temos:
```

```
# mulher = 50.5%
```

```
# homem = 49.5%
```

```
# $bmi: distribuição simétrica.
```

```
hist(Despesas$bmi, xlab = "bmi (Massa corporal)")
```



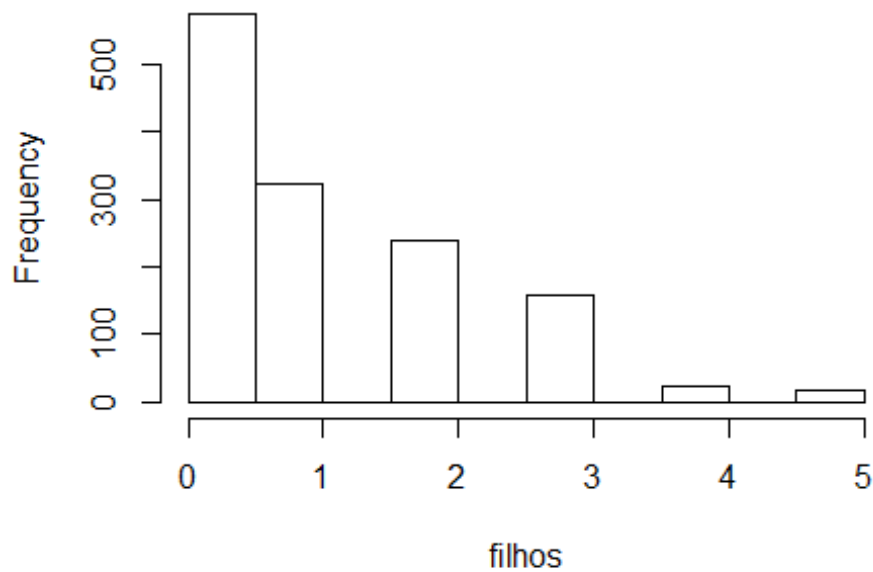
```
count(Despesas, bmi)

## # A tibble: 275 x 2
##   bmi      n
##   <dbl> <int>
## 1  16         1
## 2  16.8        2
## 3  17.2        1
## 4  17.3         3
## 5  17.4         2
## 6  17.5         1
## 7  17.7         1
## 8  17.8         2
## 9  17.9         1
## 10 18          1
## # ... with 265 more rows
```

Os dados de bmi (massa corporal) fornecidos tendem a centralidade e os valores são mais comuns entre 25 a 35 bmi.

```
# $filhos: distribuição enviesada a direita.
hist(Despesas$filhos, xlab = "filhos")
```

Histogram of Despesas\$filhos



```
table(Despesas$filhos)
```

```
##  
##    0    1    2    3    4    5  
## 574 324 240 157  25  18
```

Os dados fornecidos tendem a ter um enviesamento a direita, o que indica que a distribuição dos dados concentra-se em torno dos valores mínimos. Em termos de porcentagem temos:

0 filhos -> $574 / 1338 = 42,9\%$

1 filhos -> $324 / 1338 = 24,2\%$

2 filhos -> $240 / 1338 = 17,9\%$

3 filhos -> $157 / 1338 = 11,8\%$

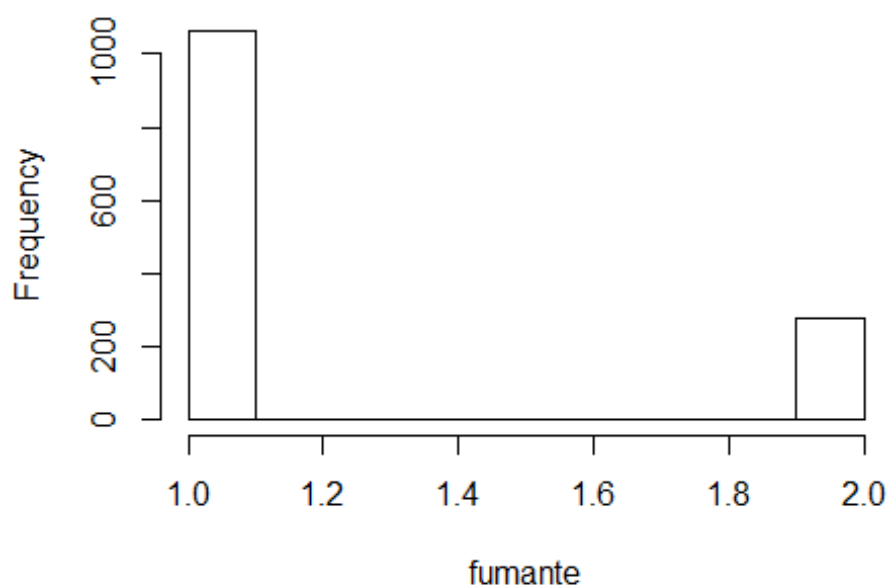
4 filhos -> $25 / 1338 = 1,9\%$

5 filhos -> $18 / 1338 = 1,3\%$

\$fumante: distribuição é dicotômica, mas desequilibrada.

```
hist(Despesas$fumante, xlab = "fumante")
```

Histogram of Despesas\$fumante



```
table(Despesas$fumante)
```

```
##  
##      1      2  
## 1064   274
```

```
round(prop.table(table(Despesas$fumante)) * 100, digits = 1)
```

```
##  
##      1      2  
## 79.5  20.5
```

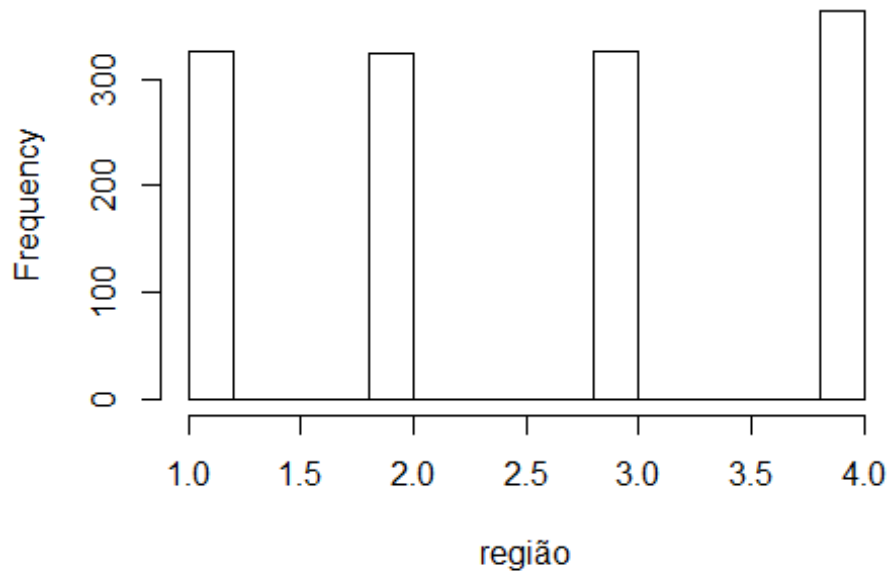
Em termos de porcentagem temos:

Não Fumante -> $1064 / 1338 = 79,5\%$

Fumante -> $274 / 1338 = 20,5\%$

```
# $regiao: distribuição uniforme  
hist(Despesas$regiao, xlab = "região")
```

Histogram of Despesas\$regiao



```
table(Despesas$regiao)
```

```
##  
##  1    2    3    4  
## 325 324 325 364
```

Normalização das variáveis:

```
# Criando um função de normalização
```

```
Normalizar <- function(x) {  
  return ((x - min(x)) / (max(x) - min(x)))  
}
```

```
Dados_Norm_2 <- as.data.frame(lapply(Dados_Norm[1:6], Normalizar))  
str(Despesas)
```

```
## 'data.frame':  1338 obs. of  7 variables:  
## $ idade  : int  19 18 28 33 32 31 46 37 37 60 ...  
## $ sexo   : num  2 1 1 1 1 2 2 2 1 2 ...  
## $ bmi    : num  27.9 33.8 33 22.7 28.9 25.7 33.4 27.7 29.8 25.8 ...  
## $ filhos : int  0 1 3 0 0 0 1 3 2 0 ...  
## $ fumante: num  2 1 1 1 1 1 1 1 1 1 ...  
## $ regiao : num  3 4 4 1 1 4 4 1 2 1 ...  
## $ gastos : num  16885 1726 4449 21984 3867 ...
```

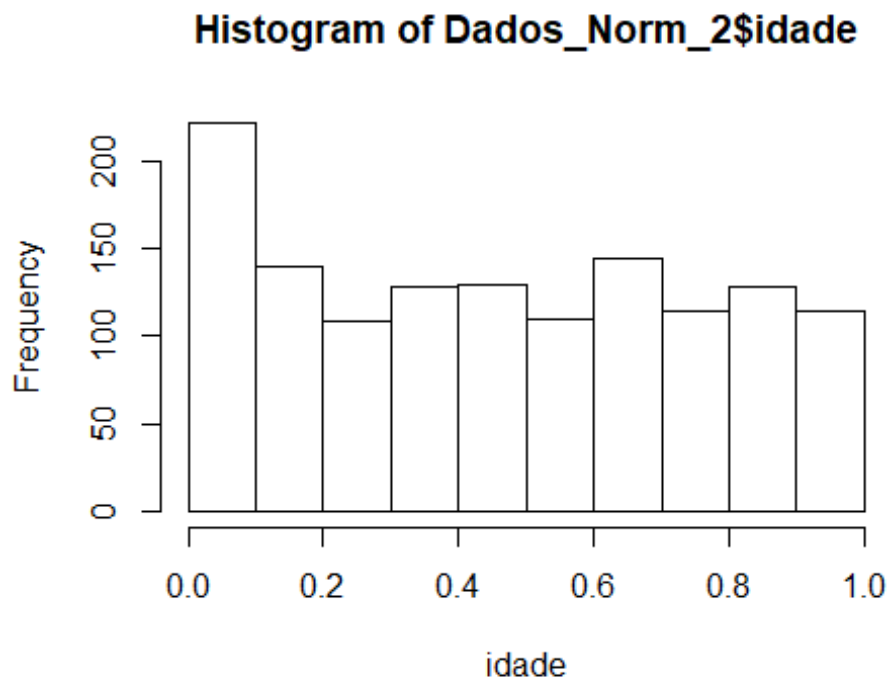
```
str(Dados_Norm_2)
```

```
## 'data.frame':  1338 obs. of  6 variables:  
## $ idade  : num  0.0217 0 0.2174 0.3261 0.3043 ...
```

```
## $ sexo : num 1 0 0 0 0 1 1 1 0 1 ...
## $ bmi : num 0.321 0.48 0.458 0.181 0.348 ...
## $ filhos : num 0 0.2 0.6 0 0 0 0.2 0.6 0.4 0 ...
## $ fumante: num 1 0 0 0 0 0 0 0 0 0 ...
## $ regioao : num 0.667 1 1 0 0 ...
```

Gráfico e Tabela de contingência para analisar a distribuição das outras variáveis após a normalização.

```
# $idade: distribuição uniforme com um enviesamento a esquerda
hist(Dados_Norm_2$idade, xlab = "idade")
```

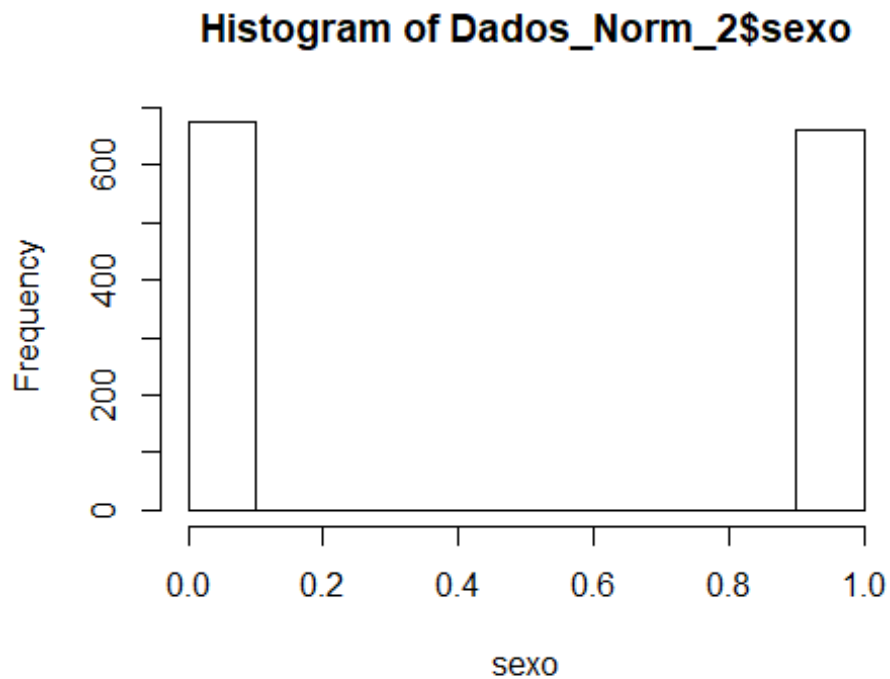


```
count(Dados_Norm_2, idade)
```

```
## # A tibble: 47 x 2
##   idade      n
##   <dbl> <int>
## 1 0         69
## 2 0.0217    68
## 3 0.0435    29
## 4 0.0652    28
## 5 0.0870    28
## 6 0.109     28
## 7 0.130     28
## 8 0.152     28
## 9 0.174     28
```

```
## 10 0.196      28
## # ... with 37 more rows

# $sexo: distribuição Dicotômica.
hist(Dados_Norm_2$sexo, xlab = "sexo")
```



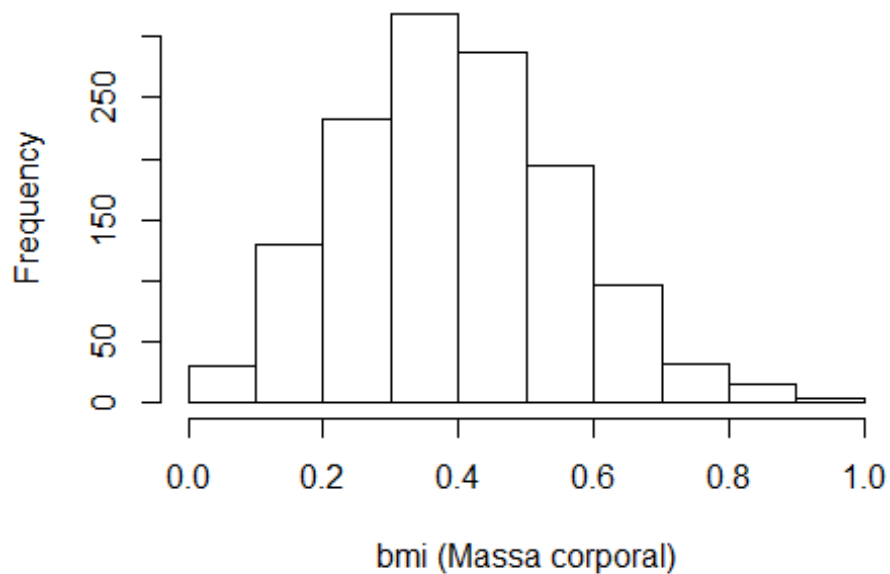
```
round(prop.table(table(Dados_Norm_2$sexo)) * 100, digits = 1)

##
##      0      1
## 50.5 49.5

# Em termos de porcentagem temos:
# mulher = 50.5%
# homem = 49.5%

# $bmi: distribuição simétrica.
hist(Dados_Norm_2$bmi, xlab = "bmi (Massa corporal)")
```


Histogram of Dados_Norm_2\$bmi



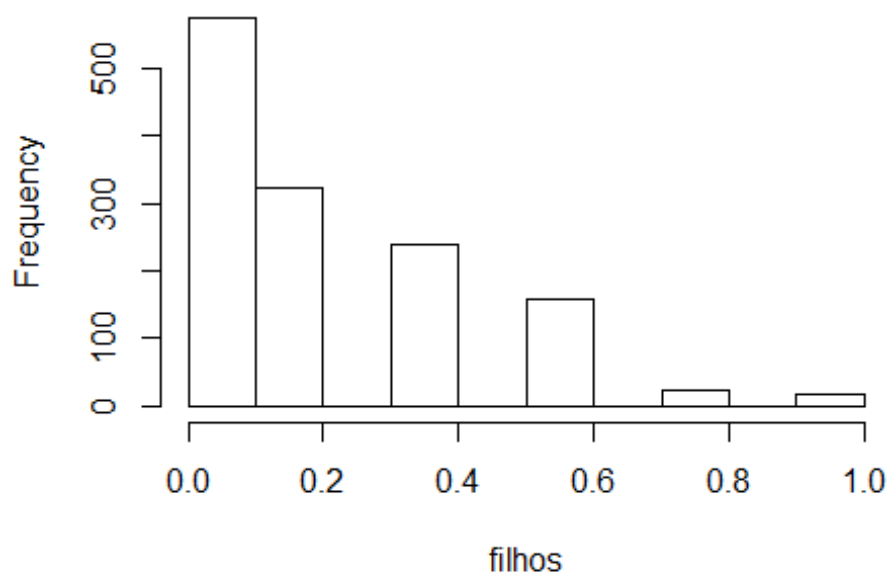
```
count(Dados_Norm_2, bmi)

## # A tibble: 275 x 2
##   bmi      n
##   <dbl> <int>
## 1 0         1
## 2 0.0216     2
## 3 0.0323     1
## 4 0.0350     3
## 5 0.0377     2
## 6 0.0404     1
## 7 0.0458     1
## 8 0.0485     2
## 9 0.0512     1
## 10 0.0539     1
## # ... with 265 more rows
```

Os dados de bmi (massa corporal) fornecidos tendem a centralidade.

```
# $filhos: distribuição enviesada a direita.
hist(Dados_Norm_2$filhos, xlab = "filhos")
```

Histogram of Dados_Norm_2\$filhos



```
table(Dados_Norm_2$filhos)
```

```
##
```

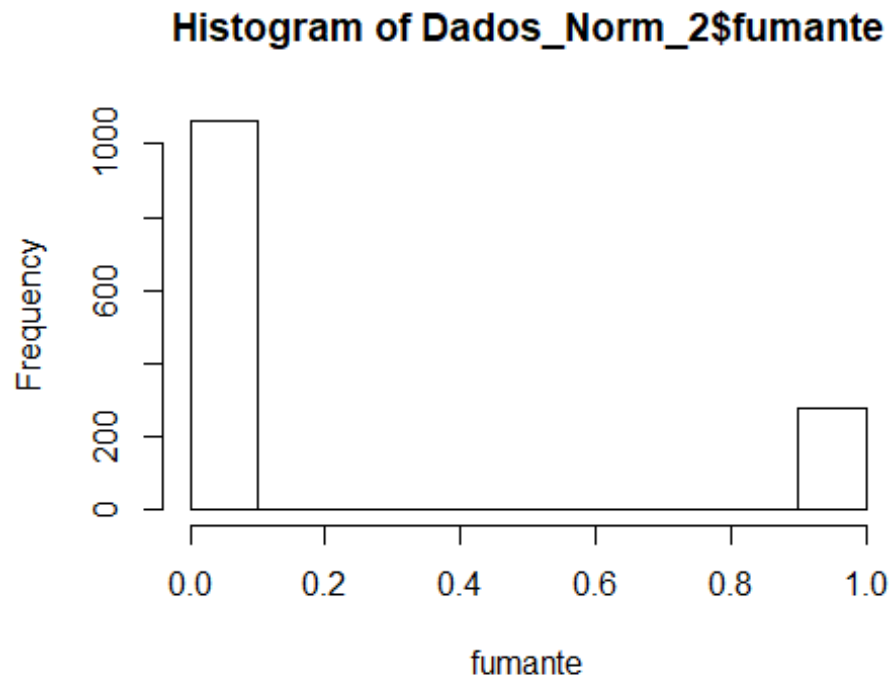
```
##  0 0.2 0.4 0.6 0.8  1
```

```
## 574 324 240 157  25  18
```

Os dados fornecidos tendem a ter um enviesamento a direita, o que indica que a distribuição dos dados concentra-se em torno dos valores mínimos.

```
# $fumante: distribuição é dicotômica, mas desequilibrada.
```

```
hist(Dados_Norm_2$fumante, xlab = "fumante")
```



```
table(Dados_Norm_2$fumante)
```

```
##  
##      0      1  
## 1064   274
```

```
round(prop.table(table(Dados_Norm_2$fumante)) * 100, digits = 1)
```

```
##  
##      0      1  
## 79.5  20.5
```

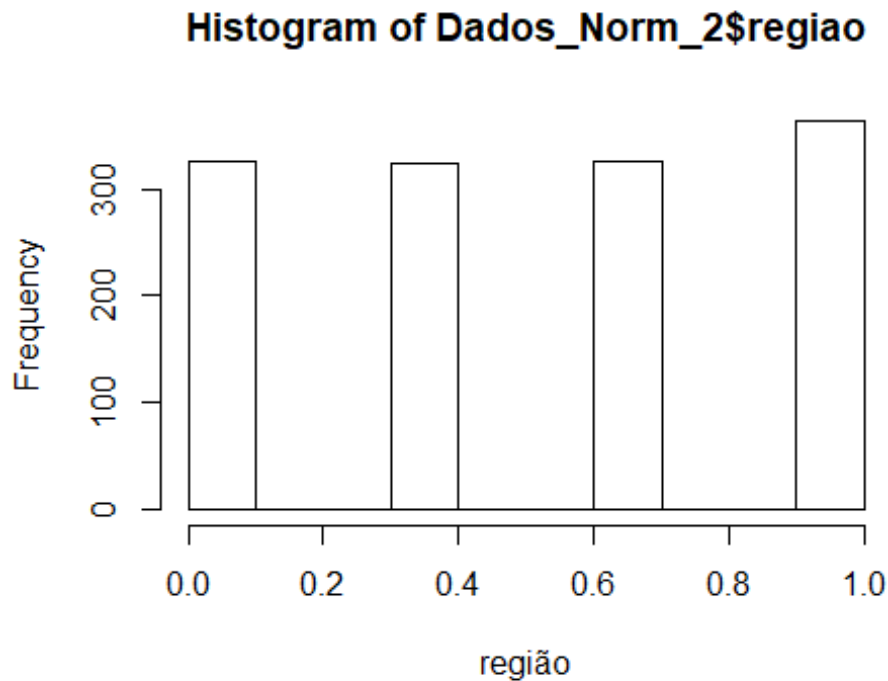
Em termos de porcentagem temos:

Não Fumante -> $1064 / 1338 = 79,5\%$

Fumante -> $274 / 1338 = 20,5\%$

```
# $regiao: distribuição uniforme
```

```
hist(Dados_Norm_2$regiao, xlab = "região")
```



```
table(Dados_Norm_2$regiao)

##
##           0 0.333333333333333 0.666666666666667
1
##           325           324           325           36
4
```

Não observei diferença na distribuição dos dados após a normalização das variáveis. Sendo assim irei considerar a normalização apenas para a variável \$gastos.

```
str(Dados_Norm)

## 'data.frame':   1338 obs. of  7 variables:
## $ idade  : int  19 18 28 33 32 31 46 37 37 60 ...
## $ sexo   : num  2 1 1 1 1 2 2 2 1 2 ...
## $ bmi    : num  27.9 33.8 33 22.7 28.9 25.7 33.4 27.7 29.8 25.8 ...
## $ filhos : int  0 1 3 0 0 0 1 3 2 0 ...
## $ fumante: num  2 1 1 1 1 1 1 1 1 1 ...
## $ regiao : num  3 4 4 1 1 4 4 1 2 1 ...
## $ gastos : num  9.73 7.45 8.4 10 8.26 ...
```

Explorando relacionamento entre as variáveis: Matriz de Correlação

```
# Visualizando relacionamento entre as variáveis:
#install.packages("Hmisc")
library(Hmisc)
```

```

## Loading required package: survival

##
## Attaching package: 'survival'

## The following object is masked from 'package:caret':
##
##   cluster

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

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

## The following objects are masked from 'package:dplyr':
##
##   src, summarize

## The following objects are masked from 'package:base':
##
##   format.pval, units

m <- rcorr(as.matrix(Dados_Norm)) #coeficientes de correlacao, n e valor
p
m

##          idade  sexo   bmi  filhos  fumante  regiao  gastos
## idade      1.00  0.02  0.11   0.04   -0.03  -0.01   0.53
## sexo       0.02  1.00 -0.05  -0.02   -0.08  -0.02  -0.01
## bmi        0.11 -0.05  1.00   0.01    0.00   0.26   0.13
## filhos     0.04 -0.02  0.01   1.00    0.01  -0.02   0.16
## fumante    -0.03 -0.08  0.00   0.01    1.00   0.05   0.67
## regiao     -0.01 -0.02  0.26  -0.02    0.05   1.00   0.00
## gastos     0.53 -0.01  0.13   0.16    0.67   0.00   1.00
##
## n= 1338
##
##
## P
##          idade  sexo   bmi   filhos  fumante  regiao  gastos
## idade              0.4459 0.0000 0.1205 0.3605  0.8489 0.0000
## sexo      0.4459              0.0899 0.5305 0.0053  0.5557 0.8369
## bmi       0.0000 0.0899              0.6440 0.8847  0.0000 0.0000
## filhos    0.1205 0.5305 0.6440              0.7792  0.4816 0.0000
## fumante   0.3605 0.0053 0.8847 0.7792              0.0486 0.0000
## regiao    0.8489 0.5557 0.0000 0.4816 0.0486              0.9011
## gastos    0.0000 0.8369 0.0000 0.0000 0.0000 0.9011

```

```
m$r #coeficiente de correlação
```

```
##          idade          sexo          bmi          filhos          fumante
## idade      1.00000000  0.020855872  0.109341015  0.04246900 -0.025018752
## sexo      0.02085587  1.000000000 -0.046380206 -0.01716298 -0.076184817
## bmi       0.10934101 -0.046380206  1.000000000  0.01264471  0.003968054
## filhos    0.04246900 -0.017162978  0.012644707  1.000000000  0.007673120
## fumante   -0.02501875 -0.076184817  0.003968054  0.00767312  1.000000000
## regioao  -0.00521169 -0.016120988  0.261848704 -0.01925722  0.053929632
## gastos    0.52783405 -0.005631878  0.132787550  0.16133635  0.665505714
##          regioao          gastos
## idade   -0.005211690  0.527834047
## sexo    -0.016120988 -0.005631878
## bmi      0.261848704  0.132787550
## filhos  -0.019257217  0.161336351
## fumante  0.053929632  0.665505714
## regioao  1.000000000  0.003400914
## gastos   0.003400914  1.000000000
```

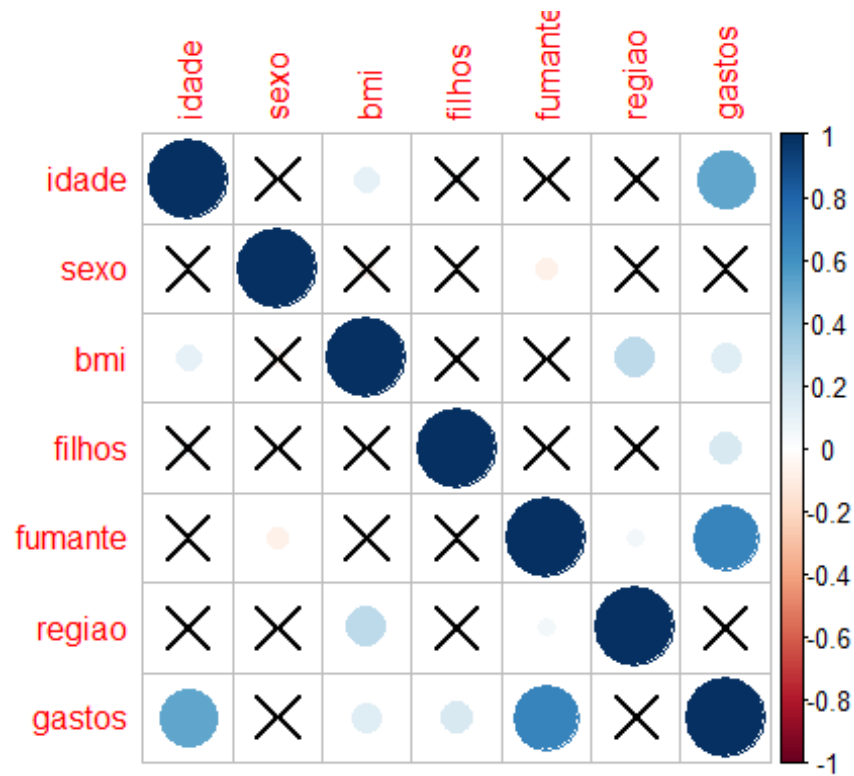
```
m$p #valor p
```

```
##          idade          sexo          bmi          filhos          fumante
## idade      NA 0.445910681 6.127822e-05 1.204928e-01 0.360485290
## sexo      4.459107e-01 NA 8.991357e-02 5.304898e-01 0.005300369
## bmi       6.127822e-05 0.089913570 NA 6.440004e-01 0.884702101
## filhos    1.204928e-01 0.530489770 6.440004e-01 NA 0.779159566
## fumante   3.604853e-01 0.005300369 8.847021e-01 7.791596e-01 NA
## regioao   8.489488e-01 0.555746126 0.000000e+00 4.815508e-01 0.048579193
## gastos    0.000000e+00 0.836934827 1.092208e-06 2.940685e-09 0.000000000
##          regioao          gastos
## idade    0.84894879 0.000000e+00
## sexo     0.55574613 8.369348e-01
## bmi      0.00000000 1.092208e-06
## filhos   0.48155079 2.940685e-09
## fumante  0.04857919 0.000000e+00
## regioao   NA 9.010896e-01
## gastos   0.90108959 NA
```

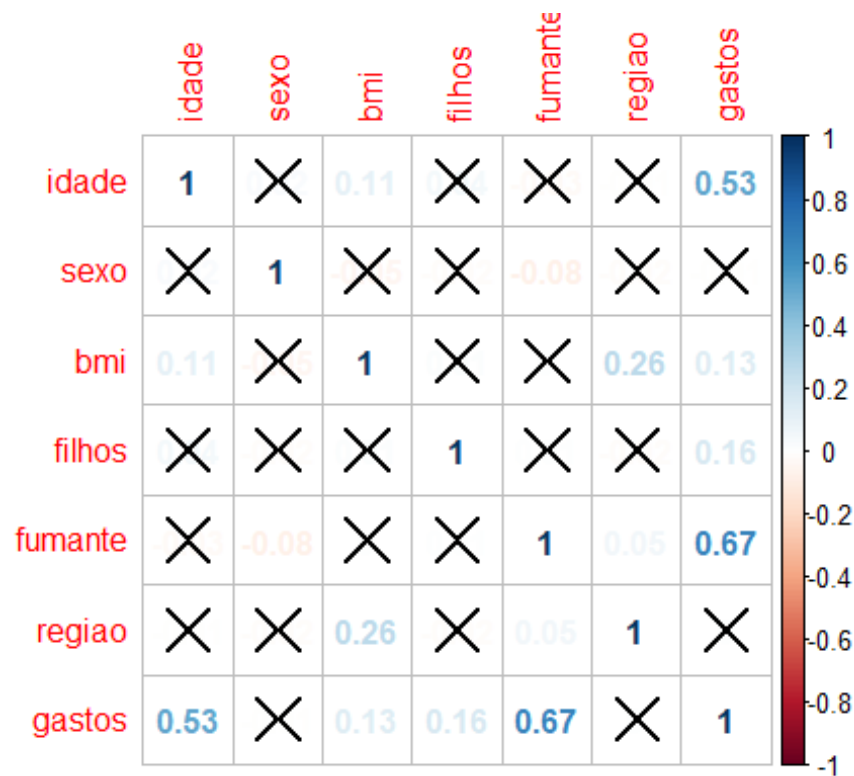
```
m$n #variáveis
```

```
##          idade sexo  bmi  filhos  fumante  regioao  gastos
## idade    1338 1338 1338   1338   1338   1338   1338
## sexo     1338 1338 1338   1338   1338   1338   1338
## bmi      1338 1338 1338   1338   1338   1338   1338
## filhos   1338 1338 1338   1338   1338   1338   1338
## fumante  1338 1338 1338   1338   1338   1338   1338
## regioao  1338 1338 1338   1338   1338   1338   1338
## gastos   1338 1338 1338   1338   1338   1338   1338
```

```
corrplot(m$r, p.mat = m$p, sig.level = 0.05, method = "circle")
```



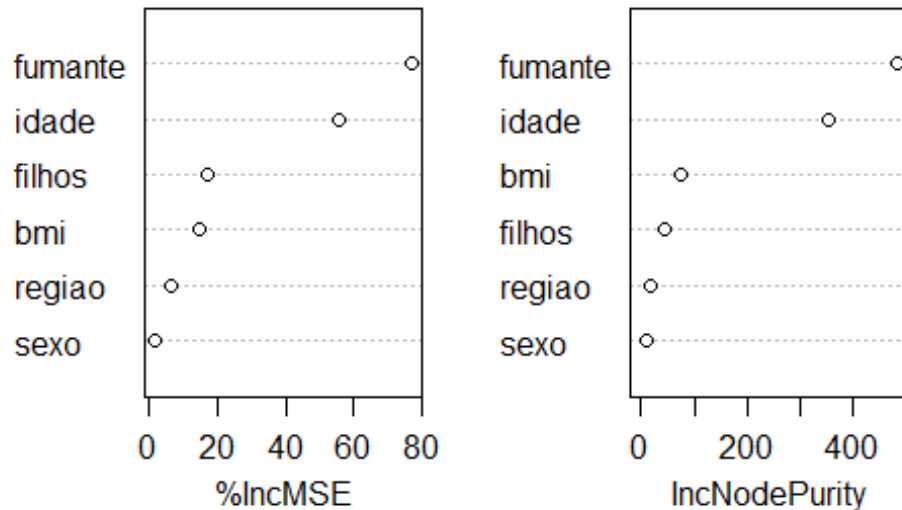
```
corrplot(m$r, p.mat = m$p, sig.level = 0.05, method = "number")
```



O objetivo do estudo da correlação é determinar (mensurar) o grau de relacionamento entre duas variáveis. Existem algumas associações interessantes. Significância = 0.05. Pode-se destacar que a correlação entre as variáveis é positiva e moderada. Por ordem de força temos a fumante, idade, filhos e bmi. Estas associações implicam que, à medida que elas aumentam, o custo esperado do seguro saúde sobe! Irei considerar apenas as variáveis que causam impacto.

```
# Vamos Criar um modelo para compararmos com os resultados obtidos acima  
para identificar os atributos com maior importância para o modelo prediti  
vo  
require(randomForest)  
  
## Loading required package: randomForest  
  
## randomForest 4.6-14  
  
## Type rfNews() to see new features/changes/bug fixes.  
  
##  
## Attaching package: 'randomForest'  
  
## The following object is masked from 'package:psych':  
##  
##     outlier  
  
## The following object is masked from 'package:ggplot2':  
##  
##     margin  
  
## The following object is masked from 'package:dplyr':  
##  
##     combine  
  
# Avaliando a importância de todas as variáveis  
Imp_Var <- randomForest(gastos ~ ., data = Dados_Norm, ntree = 100, nodes  
ize = 10,  
                        importance = TRUE)  
varImpPlot(Imp_Var)
```


Imp_Var



Confirmamos que as variáveis sexo e região tem pouca significância. Portanto elas serão removidas.

```
Dados_Norm_3 <- Dados_Norm
Dados_Norm_3 <- Dados_Norm_3[, -c(2,6)]
Data_Cor <- cor(Dados_Norm_3)
Data_Cor
```

	idade	bmi	filhos	fumante	gastos
idade	1.00000000	0.109341015	0.04246900	-0.025018752	0.5278340
bmi	0.10934101	1.000000000	0.01264471	0.003968054	0.1327875
filhos	0.04246900	0.012644707	1.00000000	0.007673120	0.1613364
fumante	-0.02501875	0.003968054	0.00767312	1.000000000	0.6655057
gastos	0.52783405	0.132787550	0.16133635	0.665505714	1.0000000

```
# Visualizando relacionamento entre as variáveis:
m2 <- rcorr(as.matrix(Dados_Norm_3)) #coeficientes de correlação, n e val
or p
m2
```

	idade	bmi	filhos	fumante	gastos
idade	1.00	0.11	0.04	-0.03	0.53
bmi	0.11	1.00	0.01	0.00	0.13
filhos	0.04	0.01	1.00	0.01	0.16
fumante	-0.03	0.00	0.01	1.00	0.67
gastos	0.53	0.13	0.16	0.67	1.00

```
## n= 1338
```

```
##
```

```
##
```

```
## P
```

```
##      idade  bmi    filhos fumante gastos
## idade      0.0000 0.1205 0.3605 0.0000
## bmi        0.0000      0.6440 0.8847 0.0000
## filhos    0.1205 0.6440      0.7792 0.0000
## fumante   0.3605 0.8847 0.7792      0.0000
## gastos    0.0000 0.0000 0.0000 0.0000
```

```
m2$r #coeficiente de correlação
```

```
##      idade      bmi    filhos    fumante    gastos
## idade  1.00000000 0.109341015 0.04246900 -0.025018752 0.5278340
## bmi    0.10934101 1.000000000 0.01264471 0.003968054 0.1327875
## filhos 0.04246900 0.012644707 1.000000000 0.007673120 0.1613364
## fumante -0.02501875 0.003968054 0.00767312 1.000000000 0.6655057
## gastos 0.52783405 0.132787550 0.16133635 0.665505714 1.0000000
```

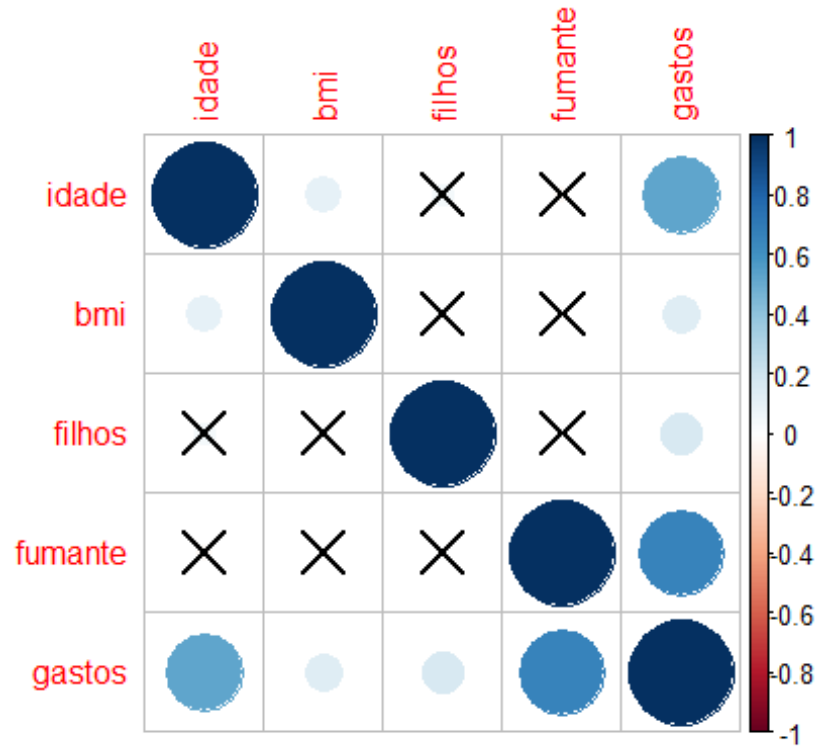
```
m2$P #valor p
```

```
##      idade      bmi    filhos    fumante    gastos
## idade      NA 6.127822e-05 1.204928e-01 0.3604853 0.000000e+00
## bmi        6.127822e-05      NA 6.440004e-01 0.8847021 1.092208e-06
## filhos    1.204928e-01 6.440004e-01      NA 0.7791596 2.940685e-09
## fumante   3.604853e-01 8.847021e-01 7.791596e-01      NA 0.000000e+00
## gastos    0.000000e+00 1.092208e-06 2.940685e-09 0.0000000      NA
```

```
m2$n #variáveis
```

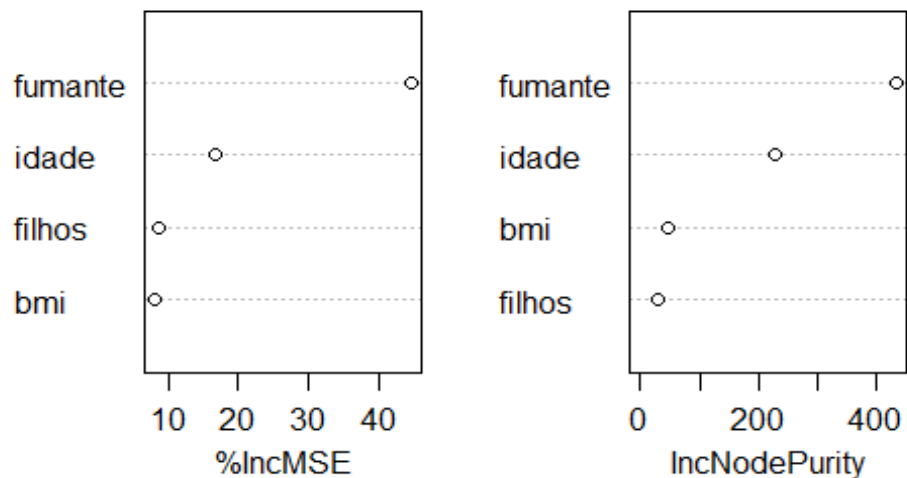
```
##      idade  bmi  filhos  fumante  gastos
## idade  1338 1338   1338   1338   1338
## bmi    1338 1338   1338   1338   1338
## filhos 1338 1338   1338   1338   1338
## fumante 1338 1338   1338   1338   1338
## gastos 1338 1338   1338   1338   1338
```

```
corrplot(m2$r, p.mat = m2$P, sig.level = 0.005, method = "circle")
```



```
# Avaliando a importância de todas as variáveis
Imp_Var_2 <- randomForest(gastos ~ ., data = Dados_Norm_3, ntree = 100, n
odesize = 10,
                          importance = TRUE)
varImpPlot(Imp_Var_2)
```

Imp_Var_2



Etapa 3 - Modelagem

Treinando o Modelo. No modelo é usado a Regressão Linear Múltipla. Formula da Regressão Linear múltiplas Estimada: $y = a + b_0x_1 + b_1x_2 + \dots + b_nx_n$

Criando amostras randômicas:

```
# Criando amostras randômicas com o caret
set.seed(1)
Amostra <- createDataPartition(y = Dados_Norm_3$gastos, p = 0.7, list = F
ALSE)

# Criando dados de treino e de teste. Divide de acordo com A AMOSTRA (+ e
-)
Dados_Treino <- Dados_Norm_3[Amostra,]
Dados_Testes <- Dados_Norm_3[-Amostra,]
#Verificando a amostra
Dados_Treino
```

##	idade	bmi	filhos	fumante	gastos
## 1	19	27.9	0	2	9.734176
## 4	33	22.7	0	1	9.998092
## 5	32	28.9	0	1	8.260198
## 7	46	33.4	1	1	9.016827
## 8	37	27.7	3	1	8.893094
## 10	60	25.8	0	1	10.272397
## 11	25	26.2	0	1	7.908872
## 12	62	26.3	0	2	10.233105
## 13	23	34.4	0	1	7.510343
## 14	56	39.8	0	1	9.313864
## 15	27	42.1	0	2	10.586881
## 16	19	24.6	1	1	7.516020
## 17	52	30.8	1	1	9.287055
## 18	23	23.8	0	1	7.781209
## 19	56	40.3	0	1	9.268835
## 20	30	35.3	0	2	10.514271
## 24	34	31.9	1	2	10.537465
## 26	59	27.7	3	1	9.546893
## 29	23	17.4	1	1	7.928474
## 30	31	36.3	2	2	10.563879
## 31	22	35.6	0	2	10.479696
## 33	19	28.6	5	1	8.452719
## 34	63	28.3	0	1	9.530255
## 35	28	36.4	1	2	10.843389
## 36	19	20.4	0	1	7.393528
## 37	62	33.0	3	1	9.655807
## 38	26	20.8	0	1	7.741664
## 39	35	36.7	1	2	10.590976
## 40	60	39.9	0	2	10.782561
## 41	24	26.6	0	1	8.021604

## 43	41 21.8	1	1	8.743927
## 44	37 30.8	2	1	8.750487
## 45	38 37.1	1	1	8.712706
## 46	55 37.3	0	1	9.934515
## 47	18 38.7	2	1	8.129576
## 51	18 35.6	0	1	7.701259
## 52	21 33.6	2	1	8.183071
## 53	48 28.0	1	2	10.067657
## 54	36 34.4	0	2	10.538544
## 55	40 28.7	3	1	8.994629
## 57	58 31.8	2	1	9.518367
## 58	18 31.7	2	2	10.442993
## 60	34 37.3	2	1	8.697767
## 62	25 33.7	4	1	8.412868
## 63	64 24.7	1	1	10.314491
## 65	20 22.4	0	2	9.596401
## 66	19 28.9	0	1	7.463484
## 67	61 39.1	2	1	9.563464
## 69	40 36.2	0	1	8.686109
## 72	31 28.5	5	1	8.824598
## 73	53 28.1	3	1	9.370904
## 74	58 32.0	1	1	9.388205
## 75	44 27.4	2	1	8.952457
## 76	57 34.0	0	1	9.337560
## 77	29 29.6	1	1	8.280815
## 82	45 38.3	0	1	8.979075
## 84	48 41.2	4	1	9.308706
## 85	37 34.8	2	2	10.592539
## 87	57 31.2	0	2	10.682329
## 89	46 27.7	0	1	8.990525
## 91	21 39.5	0	1	7.614297
## 92	53 24.8	1	1	9.300376
## 93	59 29.8	3	2	10.315098
## 94	35 34.8	2	1	8.653298
## 95	64 31.3	2	2	10.764077
## 96	28 37.6	1	1	8.234002
## 97	54 30.8	3	1	9.401400
## 100	38 19.3	0	2	9.669074
## 101	41 31.6	0	1	8.730065
## 102	30 25.5	0	1	8.201136
## 103	18 30.1	0	1	9.968566
## 105	34 27.5	1	1	8.517963
## 106	20 28.0	1	2	9.773401
## 107	19 28.4	1	1	7.754276
## 109	29 27.9	0	1	7.961063
## 111	54 33.6	1	1	9.289637
## 113	37 30.8	0	1	8.443925
## 116	60 28.6	0	1	10.317582
## 117	58 49.1	0	1	9.339730
## 119	49 27.2	0	1	9.059672

## 120	37 23.4	2	1	8.807835
## 122	18 23.8	0	1	7.441684
## 123	20 29.0	0	1	7.722004
## 125	47 33.9	3	1	9.221776
## 126	26 28.8	0	1	8.127227
## 127	19 28.3	0	2	9.745727
## 128	52 37.4	0	1	9.173110
## 129	32 17.8	2	2	10.396175
## 132	61 22.0	0	1	9.519027
## 133	53 35.9	2	1	9.320411
## 134	19 25.6	0	1	7.397905
## 135	20 28.8	0	1	7.806782
## 139	54 31.9	3	1	10.215474
## 140	22 36.0	0	1	7.680974
## 141	34 22.4	2	1	10.217418
## 142	26 32.5	1	1	8.157815
## 143	34 25.3	2	2	9.850746
## 144	29 29.7	2	1	9.806860
## 145	30 28.7	3	2	9.940108
## 146	29 38.8	3	1	8.544470
## 147	46 30.5	3	2	10.614488
## 149	53 37.4	1	1	9.301979
## 150	19 28.4	1	1	7.518889
## 151	35 24.1	1	1	8.541929
## 152	48 29.7	0	1	8.960550
## 154	42 23.4	0	2	9.901723
## 157	48 24.4	0	2	9.962873
## 158	18 25.2	0	2	9.649768
## 159	30 35.5	0	2	10.517328
## 160	50 27.8	3	1	9.890877
## 161	42 26.6	0	2	9.968747
## 162	18 36.9	0	2	10.495418
## 163	54 39.6	1	1	9.254410
## 164	32 29.8	2	1	8.547166
## 165	37 29.6	0	1	8.522807
## 166	47 28.2	4	1	9.250243
## 167	20 37.0	5	1	8.482732
## 168	32 33.2	3	1	8.720754
## 169	19 31.8	1	1	7.908122
## 171	63 41.5	0	1	9.503412
## 172	49 30.3	0	1	9.001676
## 173	18 16.0	0	1	7.435320
## 174	35 34.8	1	1	8.565231
## 176	63 37.7	0	2	10.795986
## 177	38 27.8	2	1	8.772744
## 179	46 28.9	2	1	9.085149
## 180	41 33.2	3	1	9.052316
## 182	18 38.3	0	1	7.397451
## 183	22 20.0	3	1	8.295404
## 185	44 30.7	2	1	8.953049

## 188	30 30.9	3	1	8.580290
## 189	41 32.2	1	1	8.821136
## 190	29 32.1	2	1	8.501657
## 192	36 26.2	0	1	8.493693
## 194	56 26.6	1	1	9.396350
## 198	45 28.6	2	1	9.049799
## 202	48 32.2	1	1	9.090560
## 203	60 24.0	0	1	9.473643
## 205	46 22.3	0	1	8.874463
## 206	28 28.9	1	1	8.375109
## 207	59 26.4	0	1	9.371038
## 208	35 27.7	2	2	9.951520
## 210	40 41.2	1	1	8.796356
## 211	20 33.0	1	1	7.590887
## 212	40 30.9	4	1	9.007333
## 213	24 28.5	2	1	8.171232
## 214	34 26.7	1	1	8.517749
## 217	53 26.6	0	1	9.245287
## 219	26 29.9	1	1	8.129464
## 220	24 23.2	0	1	10.129897
## 221	34 33.7	1	1	8.519684
## 222	53 33.3	0	1	9.265291
## 223	32 30.8	3	1	8.566654
## 224	19 34.8	0	2	10.456787
## 225	42 24.6	0	2	9.878966
## 226	55 33.9	3	1	9.391592
## 227	28 38.1	0	1	7.897111
## 228	58 41.9	0	1	10.095237
## 229	41 31.6	1	1	8.903568
## 231	42 36.2	1	1	8.915115
## 232	59 27.8	3	1	9.546905
## 233	19 17.8	0	1	7.454598
## 235	39 24.5	2	1	8.811383
## 236	40 22.2	2	2	9.875308
## 237	18 26.7	0	1	7.387567
## 238	31 38.4	2	1	8.403624
## 242	33 22.1	1	1	8.585612
## 243	55 26.8	1	1	10.467668
## 244	40 35.3	3	1	8.881401
## 247	60 38.1	0	1	9.445310
## 249	19 20.9	1	1	7.513213
## 250	29 29.0	1	1	8.304139
## 251	18 17.3	2	2	9.459499
## 252	63 32.2	2	2	10.764378
## 253	54 34.2	2	2	10.697854
## 254	27 30.3	3	1	8.357198
## 255	50 31.8	0	2	10.623694
## 257	56 33.6	0	2	10.690152
## 258	38 40.2	0	1	8.594336
## 259	51 24.4	4	1	9.351849

## 261	58 25.2	0	1	9.378999
## 262	20 26.8	1	2	9.745972
## 264	19 37.0	0	2	10.497350
## 267	40 19.8	1	2	9.751473
## 268	59 32.4	3	1	9.588135
## 270	49 25.8	1	1	9.135884
## 271	18 29.4	1	1	7.449754
## 273	41 37.1	2	1	8.890920
## 274	50 27.5	1	1	9.171356
## 275	25 27.6	0	1	7.833271
## 276	47 26.6	2	1	9.181513
## 278	22 24.3	0	1	7.673442
## 279	59 31.8	2	1	9.467212
## 281	40 28.1	1	2	10.013757
## 282	54 40.6	3	2	10.790333
## 284	55 32.4	1	1	9.382536
## 285	52 31.2	0	1	9.172215
## 286	46 26.6	1	1	8.954430
## 287	46 48.1	2	1	9.151962
## 288	63 26.2	0	1	9.564946
## 289	59 36.8	1	2	10.776804
## 293	25 45.5	2	2	10.648094
## 295	25 26.8	3	1	8.270302
## 296	18 23.0	0	1	7.441068
## 298	47 25.4	1	2	9.997828
## 299	31 34.4	3	2	10.564792
## 302	53 22.6	3	2	10.121553
## 303	56 37.5	2	1	9.414547
## 305	57 38.0	2	1	9.445113
## 306	29 33.3	2	1	9.875209
## 307	28 27.5	2	1	9.912332
## 308	30 33.3	1	1	8.331112
## 311	50 26.6	0	1	9.041267
## 312	19 24.7	0	1	7.460134
## 313	43 36.0	3	2	10.648385
## 314	49 35.9	0	1	9.002628
## 315	27 31.4	0	2	10.458489
## 316	52 33.3	0	1	9.182226
## 317	50 32.2	0	1	9.086506
## 319	44 27.6	0	1	8.912095
## 320	32 37.3	1	1	8.448402
## 321	34 25.3	1	1	8.495918
## 325	29 27.2	0	1	7.960704
## 327	27 23.2	1	1	8.178047
## 328	45 36.5	2	2	10.663370
## 329	64 33.8	1	2	10.777456
## 331	61 36.4	1	2	10.789681
## 333	61 31.2	0	1	9.505175
## 334	56 28.8	0	1	9.363781
## 335	43 35.7	2	1	9.859775

## 336	64 34.5	0	1	9.534075
## 337	60 25.7	0	1	9.404474
## 338	62 27.6	1	1	9.542351
## 339	50 32.3	1	2	10.643497
## 344	63 36.8	0	1	9.545515
## 345	49 41.5	4	1	9.303577
## 346	34 29.3	3	1	8.729769
## 352	50 25.6	0	1	9.097405
## 353	30 27.7	0	1	8.175885
## 354	33 35.2	0	1	9.425845
## 355	18 38.3	0	1	9.556271
## 357	46 43.9	3	1	9.098752
## 360	18 20.8	0	1	7.382442
## 361	48 32.3	2	1	9.214656
## 364	21 26.4	1	1	7.862413
## 365	21 21.9	2	1	8.064797
## 366	49 30.8	1	1	9.187926
## 367	56 32.3	3	1	9.505266
## 368	42 25.0	2	1	8.989327
## 369	44 32.0	2	1	9.001626
## 370	18 30.4	3	1	8.155325
## 371	61 21.1	0	1	9.504132
## 372	57 22.2	0	1	9.395100
## 373	42 33.2	1	1	8.941077
## 375	20 33.3	0	1	7.238159
## 376	23 28.3	0	2	9.800012
## 377	39 24.9	3	2	9.983219
## 378	24 40.2	0	2	10.548658
## 380	62 31.5	1	1	10.203628
## 381	27 18.0	2	2	9.616244
## 382	55 30.7	0	2	10.652630
## 383	55 33.0	0	1	9.941818
## 385	44 22.1	2	1	9.024317
## 387	58 39.1	0	1	9.380624
## 388	50 25.4	2	1	10.318396
## 389	26 22.6	0	1	8.063636
## 390	24 30.2	3	1	8.437734
## 392	19 37.4	0	1	7.667659
## 393	48 31.4	1	1	9.100979
## 394	49 31.4	1	1	9.136709
## 395	46 32.3	2	1	9.149636
## 398	21 31.0	0	1	9.716344
## 399	64 25.6	2	1	9.615034
## 400	18 38.2	0	1	7.397359
## 401	51 20.6	0	1	9.133978
## 402	47 47.5	1	1	8.997632
## 404	49 32.3	3	1	9.236930
## 405	31 20.4	0	1	8.089544
## 406	52 38.4	2	1	9.341097
## 408	47 23.6	1	1	9.052478

## 409	38 21.1	3	1	8.802753
## 410	32 30.0	1	1	8.312491
## 411	19 17.5	0	1	7.391008
## 413	26 17.2	2	2	9.578840
## 415	19 35.2	0	1	7.666175
## 417	52 34.1	0	1	9.120520
## 418	36 22.6	2	2	9.831361
## 419	64 39.2	1	1	9.576252
## 420	63 27.0	0	2	10.273342
## 423	40 32.8	1	2	10.574525
## 424	25 30.6	0	1	7.911104
## 425	48 30.2	2	1	9.101455
## 428	18 29.2	0	1	8.898875
## 429	21 16.8	1	1	8.060685
## 430	27 30.4	3	1	9.841865
## 431	19 33.1	0	1	10.046850
## 432	29 20.2	2	1	8.498298
## 434	60 30.5	0	1	9.444479
## 435	31 28.6	1	1	8.353165
## 438	35 28.9	3	1	8.687248
## 439	52 46.8	5	1	9.440859
## 440	26 29.5	0	1	7.971541
## 441	31 32.7	1	1	8.463427
## 442	33 33.5	0	2	10.520816
## 443	18 43.0	0	1	7.046995
## 446	45 33.1	0	1	8.901786
## 447	60 29.6	0	1	9.451795
## 448	56 25.7	0	1	9.346096
## 449	40 29.6	0	1	8.684560
## 451	39 29.6	4	1	8.924293
## 452	30 24.1	1	1	8.302077
## 453	24 23.4	0	1	7.585591
## 454	20 29.7	0	1	7.478469
## 455	32 46.5	2	1	8.452418
## 456	59 37.4	0	1	9.989528
## 457	55 30.1	2	1	9.382777
## 459	56 39.6	0	1	9.268742
## 460	40 33.0	3	1	8.946722
## 461	49 36.6	3	1	9.247779
## 462	42 30.0	0	2	10.005323
## 463	62 38.1	2	1	9.631043
## 465	19 25.2	0	1	7.397586
## 467	60 28.7	1	1	9.489841
## 469	28 24.3	1	1	10.055733
## 471	27 32.7	0	1	7.822861
## 472	18 30.1	0	1	7.697789
## 473	19 29.8	0	1	7.464206
## 474	47 33.3	0	1	9.946489
## 475	54 25.1	3	2	10.141807
## 476	61 28.3	1	2	10.270512

## 477	24 28.5	0	2 10.467310
## 479	21 36.9	0	1 7.335830
## 480	23 32.6	0	1 7.508946
## 481	63 41.3	3	1 9.652150
## 482	49 37.5	2	1 9.138275
## 483	18 31.4	0	1 7.391532
## 484	51 39.5	1	1 9.198275
## 486	31 31.1	0	1 8.377246
## 487	54 21.5	3	1 9.431510
## 488	19 28.7	0	1 7.134046
## 489	44 38.1	0	2 10.797229
## 490	53 31.2	1	1 9.255503
## 491	19 32.9	0	1 7.466668
## 492	61 25.1	0	1 10.106963
## 493	18 25.1	0	1 7.694607
## 494	61 43.4	0	1 9.439390
## 495	21 25.7	4	2 9.794906
## 498	45 28.7	2	1 8.990687
## 499	44 24.0	2	1 9.013242
## 500	62 39.2	0	1 9.508284
## 502	43 26.0	0	1 8.830158
## 504	19 30.3	0	2 10.390482
## 505	38 28.9	1	1 8.695236
## 506	37 30.9	3	1 8.824216
## 508	21 23.8	2	1 8.031743
## 509	24 25.3	0	1 8.020997
## 510	57 28.7	0	1 9.346206
## 513	51 22.4	0	1 9.144343
## 514	19 30.4	0	1 7.135926
## 516	58 35.7	0	1 9.338097
## 517	20 35.3	1	1 10.230064
## 518	45 30.5	2	1 9.037588
## 520	31 30.9	0	1 8.257842
## 522	32 44.2	0	1 8.292594
## 524	38 37.7	0	1 8.593713
## 525	42 26.1	1	2 10.551784
## 526	18 33.9	0	1 9.348591
## 527	19 30.6	2	1 10.088293
## 528	51 25.8	1	1 9.196346
## 529	46 39.4	1	1 9.029167
## 530	18 25.5	0	1 7.443078
## 531	57 42.1	1	2 10.792932
## 533	59 29.7	2	1 9.466988
## 534	37 36.2	0	1 9.863431
## 535	64 40.5	0	1 9.534676
## 539	46 28.1	1	1 9.015918
## 540	53 31.4	0	1 10.216327
## 541	34 38.0	3	1 8.731732
## 542	20 31.8	2	1 8.024990
## 543	63 36.3	0	1 9.538723

## 544	54 47.4	0	2 11.063045
## 545	54 30.2	0	1 9.233226
## 546	49 25.8	2	2 10.077745
## 547	28 35.4	0	1 8.092194
## 549	25 28.6	0	1 8.075153
## 550	43 46.2	0	2 10.733419
## 551	63 30.8	0	1 9.502305
## 552	32 28.9	0	1 8.287257
## 554	52 31.7	2	1 9.322567
## 556	28 23.8	2	1 8.255223
## 557	46 33.4	1	1 9.028170
## 558	34 34.2	0	1 8.277712
## 559	35 34.1	3	2 10.596220
## 562	54 32.7	0	1 9.298711
## 563	27 30.5	0	1 7.821651
## 564	50 44.8	1	1 9.111484
## 565	18 32.1	2	1 7.937825
## 566	19 30.5	0	1 7.663140
## 567	38 40.6	1	1 8.759913
## 568	41 30.6	2	1 8.889683
## 569	49 31.9	5	1 9.354692
## 570	48 40.6	2	2 10.729898
## 572	18 37.3	1	1 7.705015
## 574	62 36.9	1	1 10.361545
## 575	57 34.3	2	1 9.489793
## 576	58 27.2	0	1 9.411067
## 577	22 26.8	0	1 7.417580
## 578	31 38.1	1	2 10.977996
## 579	52 30.2	1	1 9.182407
## 580	25 23.5	0	1 8.072932
## 582	19 30.6	0	1 7.402183
## 583	39 45.4	2	1 8.757197
## 585	19 20.7	0	1 7.125138
## 586	33 28.3	1	1 8.472112
## 588	34 30.2	1	2 10.690669
## 589	61 35.9	0	1 9.520442
## 590	38 30.7	1	1 8.695646
## 591	58 29.0	0	1 9.379445
## 592	47 19.6	1	1 9.039323
## 594	21 21.9	1	2 9.639463
## 595	41 40.3	0	1 8.649827
## 598	34 33.3	1	1 8.629602
## 599	43 32.6	2	1 8.914828
## 600	52 37.5	2	1 10.418464
## 601	18 39.2	0	1 7.398199
## 602	51 31.6	0	1 9.124144
## 603	56 25.3	0	1 9.312043
## 604	64 39.1	3	1 9.685651
## 605	19 28.3	0	2 9.768182
## 607	27 25.2	0	1 8.177128

## 608	59 23.7	0	2 10.153420
## 609	28 27.0	2	1 8.397303
## 610	30 37.8	2	2 10.577489
## 611	47 29.4	1	1 9.053416
## 612	38 34.8	2	1 8.790503
## 613	18 33.2	0	1 7.699707
## 614	34 19.0	3	1 8.817748
## 615	20 33.0	0	1 7.539064
## 616	47 36.6	1	2 10.668254
## 617	56 28.6	0	1 9.363758
## 620	55 37.1	0	1 9.279273
## 621	30 31.4	1	1 8.205041
## 622	37 34.1	4	2 10.601181
## 623	49 21.3	1	1 9.125019
## 624	18 33.5	0	2 10.452124
## 626	29 26.0	0	1 8.225894
## 627	36 28.9	3	1 8.817089
## 628	33 42.5	1	1 9.334919
## 629	58 38.0	0	1 9.338377
## 632	24 29.3	0	1 7.589751
## 633	29 35.5	0	1 8.121679
## 635	51 39.7	1	1 9.147544
## 636	64 38.2	0	1 9.575742
## 638	35 38.1	2	1 10.123227
## 639	39 26.4	0	2 9.910926
## 641	33 42.4	5	1 8.804811
## 642	42 28.3	3	2 10.397801
## 643	61 33.9	0	1 9.483710
## 644	23 35.0	3	1 8.404387
## 645	43 35.3	2	1 9.841939
## 647	39 26.2	1	1 8.719901
## 650	58 33.0	0	1 9.427945
## 651	49 42.7	2	1 9.190228
## 652	53 39.6	1	1 9.266693
## 653	48 31.1	0	1 9.021673
## 654	45 36.3	2	1 9.051055
## 657	26 42.4	1	1 8.134561
## 658	27 33.2	2	1 8.308620
## 660	57 28.8	4	1 9.574595
## 663	32 31.5	1	1 8.546470
## 664	18 33.7	0	1 7.035621
## 665	64 23.0	0	2 10.204995
## 666	43 38.1	2	2 10.658680
## 668	40 32.8	2	2 10.596718
## 669	62 32.0	0	2 10.730077
## 670	40 29.8	1	1 8.779594
## 671	30 31.6	3	1 8.484170
## 674	41 31.0	0	1 8.729934
## 676	45 21.4	0	1 8.884997
## 678	60 31.4	3	2 10.739230

## 679	56 36.1	3	1	9.422508
## 680	49 23.2	2	1	9.225897
## 682	19 20.3	0	1	7.124688
## 686	53 26.4	2	1	9.327624
## 690	27 31.1	1	2	10.457559
## 691	21 27.4	0	1	7.651648
## 692	47 36.2	1	1	8.995684
## 693	20 32.4	1	1	7.767361
## 696	26 40.2	0	1	8.071297
## 697	53 32.3	2	1	10.281461
## 698	41 35.8	1	2	10.603453
## 699	56 33.7	0	1	9.303489
## 700	23 39.3	2	1	8.160693
## 702	50 44.7	0	1	9.163427
## 703	53 41.5	0	1	9.159501
## 705	47 29.5	1	1	9.097276
## 706	33 32.9	2	1	8.589521
## 707	51 38.1	0	2	10.701004
## 708	49 28.7	3	1	9.236441
## 709	31 30.5	3	1	8.718211
## 711	18 35.2	1	1	7.454454
## 712	50 23.5	2	1	9.221005
## 713	43 30.7	2	1	9.025316
## 714	20 40.5	0	1	7.593097
## 715	24 22.6	0	1	7.806900
## 716	60 28.9	0	1	9.404835
## 718	60 24.3	1	1	9.481329
## 719	51 36.7	2	1	9.291748
## 720	58 33.4	0	1	9.411779
## 721	51 40.7	0	1	9.197830
## 722	53 36.6	3	1	9.329415
## 723	62 37.4	0	1	9.471116
## 724	19 35.4	0	1	7.141443
## 725	50 27.1	1	1	9.220897
## 727	41 28.4	1	1	8.804579
## 728	29 21.8	1	2	9.720629
## 729	18 40.3	0	1	7.704181
## 731	35 24.4	3	2	9.871068
## 732	53 21.4	1	1	9.216860
## 735	59 32.1	3	1	9.547328
## 736	49 34.8	1	1	9.167839
## 737	37 38.4	0	2	10.607056
## 739	23 31.7	3	2	10.496513
## 740	29 35.5	2	2	10.705163
## 741	45 24.0	2	1	9.060038
## 743	53 34.1	0	2	10.674855
## 744	31 26.6	0	1	8.231600
## 745	50 26.4	0	1	9.085594
## 746	50 30.1	1	1	9.201336
## 748	19 21.8	0	1	7.394665

## 749	47 36.0	1	1	9.054494
## 751	37 26.4	0	2	9.880180
## 753	64 37.9	0	1	9.561739
## 754	58 22.8	0	1	9.378713
## 755	24 33.6	4	1	9.748495
## 756	31 27.6	2	1	8.523428
## 758	47 27.8	0	2	10.046090
## 759	30 37.4	3	1	8.599460
## 760	18 38.2	0	2	10.499788
## 761	22 34.6	2	1	8.275315
## 762	23 35.2	1	1	7.790266
## 763	33 27.1	1	2	9.854344
## 764	27 26.0	0	1	8.029697
## 767	47 32.3	1	1	8.995011
## 768	42 29.0	1	1	8.860874
## 771	61 36.1	3	1	10.237861
## 773	44 36.5	0	1	9.456982
## 775	41 34.2	2	1	8.890375
## 776	51 33.3	3	1	9.264875
## 777	40 32.3	2	1	8.851764
## 778	45 39.8	0	1	8.915755
## 779	35 34.3	3	1	8.688518
## 781	30 24.4	3	2	9.812425
## 783	51 36.0	1	1	9.146992
## 785	31 29.3	1	1	8.378048
## 786	35 27.7	3	1	8.766266
## 787	60 37.0	0	1	9.452594
## 788	21 36.9	0	1	7.558684
## 789	29 22.5	3	1	8.558255
## 790	62 29.9	0	1	9.507326
## 791	39 41.8	0	1	8.641573
## 793	22 23.2	0	1	7.912756
## 794	53 20.9	0	2	9.961559
## 795	39 31.9	2	1	8.883153
## 796	27 28.5	0	2	9.815243
## 797	30 44.2	2	1	8.358472
## 798	30 22.9	1	1	8.459462
## 799	58 33.1	0	1	9.379926
## 800	33 24.8	0	2	9.792809
## 801	42 26.2	1	1	8.860318
## 802	64 36.0	0	1	9.568983
## 803	21 22.3	1	1	7.651158
## 804	18 42.2	0	2	10.565987
## 807	40 41.4	1	1	10.256843
## 809	18 30.1	0	1	7.031308
## 810	25 25.8	1	1	8.104640
## 811	46 30.8	3	1	9.150051
## 813	54 21.0	2	1	9.306896
## 814	28 22.5	2	1	8.395904
## 817	24 24.2	0	1	7.952531

## 818	23 37.1	3	1	8.188022
## 822	26 17.7	0	1	7.893926
## 823	18 31.1	0	1	7.391341
## 824	44 29.8	2	1	9.014228
## 826	64 31.8	2	1	9.684652
## 827	56 31.8	2	2	10.687706
## 828	36 28.0	1	2	9.941440
## 829	41 30.8	3	2	10.586519
## 830	39 21.9	1	1	8.718907
## 831	63 33.1	0	1	9.502544
## 833	28 23.8	2	1	8.459509
## 834	58 34.4	0	1	9.371092
## 836	42 36.0	2	1	8.876311
## 837	36 31.5	0	1	8.389867
## 841	21 31.1	0	1	7.330608
## 843	23 32.8	2	2	10.491858
## 844	57 29.8	0	2	10.223174
## 847	51 34.2	1	1	9.197529
## 848	23 50.4	1	1	7.798958
## 849	27 24.1	0	1	7.997707
## 855	49 23.8	3	2	10.090254
## 856	20 29.6	0	1	7.536545
## 857	48 33.1	0	2	10.620697
## 858	25 24.1	0	2	9.668903
## 859	25 32.2	1	1	9.810174
## 860	57 28.1	0	1	9.302505
## 863	55 33.5	2	1	9.414887
## 864	36 19.9	0	1	8.604847
## 865	51 25.4	0	1	9.080513
## 866	40 29.9	2	1	8.794879
## 868	57 43.7	1	1	9.356700
## 869	61 23.7	0	1	9.482625
## 871	50 36.2	0	1	9.042847
## 872	26 29.5	1	1	8.129284
## 873	42 24.9	0	1	8.693981
## 874	43 30.1	1	1	8.831862
## 876	23 28.1	0	1	7.897337
## 877	49 27.1	1	1	10.171236
## 878	33 33.4	5	1	8.802942
## 880	37 29.5	2	1	8.750200
## 881	22 34.8	3	1	8.144116
## 882	23 27.4	1	1	7.933460
## 883	21 22.1	0	1	7.857810
## 884	51 37.1	3	2	10.741927
## 885	25 26.7	4	1	8.492486
## 886	32 28.9	1	2	9.889373
## 887	57 29.0	0	2	10.211650
## 888	36 30.0	0	1	8.570199
## 891	64 26.9	0	2	10.286400
## 893	54 24.0	0	1	9.251763

## 894	47 38.9	2	2 10.696540
## 896	61 44.0	0	1 9.477606
## 898	19 25.6	1	1 7.705965
## 899	18 40.3	0	1 7.399135
## 901	49 22.5	0	1 9.069797
## 903	26 27.3	3	1 8.447048
## 904	49 36.9	0	1 9.002797
## 906	26 29.4	2	1 8.425996
## 908	44 32.3	1	1 8.940331
## 909	63 39.8	3	1 9.627080
## 910	32 24.6	0	2 9.769745
## 911	22 28.3	1	1 7.878170
## 913	59 26.7	3	1 9.573782
## 914	44 27.5	1	1 8.939449
## 915	33 24.6	2	1 8.567413
## 916	24 34.0	0	1 7.813321
## 917	43 26.9	0	2 9.988487
## 918	45 22.9	0	2 10.465083
## 920	35 34.2	1	1 8.565074
## 921	62 25.0	0	1 9.506818
## 922	62 33.2	0	1 9.507665
## 923	38 31.0	1	1 8.610367
## 924	34 35.8	0	1 8.371106
## 925	43 23.2	0	1 8.740407
## 926	50 32.1	2	1 10.139876
## 927	19 23.4	2	1 7.977134
## 928	57 20.1	1	1 9.395352
## 931	26 46.5	1	1 7.981754
## 932	39 32.5	1	1 8.738463
## 933	46 25.8	5	1 9.219991
## 936	59 27.5	0	1 9.411960
## 937	44 29.7	2	1 10.376881
## 941	18 23.2	0	1 7.022752
## 944	19 22.6	0	1 7.395396
## 945	62 39.9	0	1 9.471386
## 947	42 35.8	2	1 8.876278
## 948	37 34.2	1	2 10.572529
## 950	25 29.7	3	2 9.900155
## 952	51 42.9	2	2 10.767703
## 953	30 28.4	1	1 8.417855
## 954	44 30.2	2	2 10.571280
## 955	34 27.8	1	2 9.903969
## 956	31 39.5	1	1 8.262489
## 958	24 26.8	1	1 9.442237
## 959	43 35.0	1	2 10.622162
## 960	48 36.7	1	1 10.256568
## 961	19 39.6	1	1 7.912097
## 962	29 25.9	0	1 8.117694
## 963	63 35.2	1	1 9.580156
## 965	52 36.8	2	1 10.183658

## 966	35 27.1	1	1 8.465129
## 967	51 24.8	2	2 10.084449
## 969	21 25.7	2	1 8.095559
## 970	39 34.3	5	1 9.059149
## 971	50 28.2	3	1 9.278246
## 972	34 23.6	0	1 8.515668
## 973	22 20.2	0	1 7.835113
## 974	19 40.5	0	1 7.472694
## 975	26 35.4	0	1 7.750451
## 978	26 29.2	1	1 7.973469
## 979	45 40.0	3	1 9.180362
## 980	36 29.9	0	1 8.494751
## 981	54 25.5	1	1 10.147104
## 982	34 21.4	0	1 8.411908
## 983	31 25.9	3	2 9.862662
## 985	20 30.1	5	1 8.500059
## 987	43 30.1	3	1 9.037183
## 989	34 34.7	0	1 8.416008
## 990	24 20.5	0	2 9.586850
## 991	26 19.8	1	1 8.125308
## 993	50 31.6	2	1 9.222113
## 995	27 20.0	3	2 9.706285
## 996	39 23.3	3	1 8.985505
## 998	63 36.9	0	1 9.538778
## 999	33 36.3	3	1 8.787487
## 1000	36 26.9	0	1 8.569372
## 1002	24 32.7	0	2 10.447927
## 1004	48 29.6	0	1 9.963273
## 1005	47 19.2	1	1 9.062715
## 1006	29 31.7	2	1 8.396920
## 1008	47 28.2	3	2 10.123234
## 1012	43 20.1	2	2 9.839895
## 1016	59 25.5	0	1 9.403024
## 1017	19 24.6	1	1 7.904423
## 1018	26 34.2	2	1 8.291028
## 1019	54 35.8	3	1 9.433107
## 1020	21 32.7	2	1 10.166580
## 1021	51 37.0	0	1 9.082347
## 1022	22 31.0	3	2 10.479977
## 1025	47 45.3	1	1 9.056007
## 1026	21 34.6	0	1 7.610942
## 1027	19 26.0	1	2 9.708135
## 1028	23 18.7	0	1 9.980235
## 1029	54 31.6	0	1 9.195270
## 1030	37 17.3	2	1 8.836080
## 1032	55 35.2	0	2 10.701531
## 1034	18 21.6	0	2 9.528639
## 1035	61 38.4	0	1 9.468856
## 1038	45 30.5	1	2 10.589749
## 1040	19 27.3	2	1 10.020989

## 1041	35 28.0	0	2	9.915162
## 1043	20 30.7	0	2	10.418579
## 1045	55 35.2	1	1	9.340848
## 1047	43 25.1	0	1	8.899055
## 1048	22 52.6	1	2	10.703276
## 1049	25 22.5	1	1	8.187068
## 1051	44 37.0	1	1	8.990085
## 1052	64 26.4	0	1	9.574606
## 1053	49 29.8	1	1	9.136482
## 1054	47 29.8	3	2	10.138935
## 1056	55 27.6	0	1	9.268090
## 1057	48 28.9	0	1	9.021299
## 1060	32 33.8	1	1	8.403514
## 1062	57 27.9	1	1	9.354806
## 1063	59 41.1	1	2	10.798968
## 1064	36 28.6	3	1	8.786945
## 1065	29 25.6	4	1	8.649776
## 1066	42 25.3	1	1	8.860144
## 1067	48 37.3	2	1	9.102554
## 1068	39 42.7	0	1	8.658243
## 1071	37 37.1	1	2	10.593422
## 1075	60 18.3	0	1	9.488297
## 1077	47 32.0	1	1	9.053844
## 1078	21 26.0	0	1	7.650768
## 1080	63 33.7	3	1	9.626517
## 1081	18 21.8	2	1	9.382952
## 1082	32 27.8	1	1	8.401648
## 1084	32 31.5	1	1	8.312994
## 1085	62 30.5	2	1	9.617122
## 1087	55 29.0	0	1	9.286963
## 1089	52 47.7	1	1	9.184911
## 1090	56 22.1	0	1	9.266446
## 1091	47 36.2	0	2	10.637683
## 1092	55 29.8	0	1	9.331366
## 1095	50 33.7	4	1	9.332500
## 1098	22 33.8	0	1	7.423348
## 1100	25 34.0	1	1	8.079345
## 1101	33 19.1	2	2	9.727722
## 1102	53 28.6	3	1	9.328427
## 1106	54 31.2	0	1	9.243672
## 1107	49 29.9	0	1	9.103663
## 1110	45 20.4	3	1	9.060141
## 1111	54 32.3	1	1	9.351181
## 1114	28 26.3	3	1	8.577756
## 1116	55 32.7	1	1	9.287995
## 1118	25 33.3	2	2	10.494729
## 1120	30 20.0	3	1	8.647068
## 1121	23 31.4	0	2	10.438994
## 1122	46 38.2	2	1	9.029677
## 1123	53 36.9	3	2	10.750673

## 1124	27 32.4	1	1	9.847102
## 1125	23 42.8	1	2	10.618988
## 1126	63 25.1	0	1	9.564836
## 1127	55 29.9	0	1	9.231577
## 1129	34 32.8	1	1	9.572088
## 1130	19 18.6	0	1	7.455241
## 1131	39 23.9	5	1	9.057457
## 1132	27 45.9	2	1	8.214311
## 1133	57 40.3	0	1	9.938325
## 1135	28 33.8	0	1	9.887020
## 1136	50 28.1	3	1	9.313401
## 1137	44 25.0	1	1	8.938993
## 1140	19 32.5	0	2	10.515932
## 1142	41 32.6	3	1	8.981496
## 1143	52 24.9	0	1	10.207953
## 1144	39 32.3	2	1	8.754331
## 1145	50 32.3	2	1	9.172680
## 1146	52 32.8	3	1	9.331594
## 1149	55 21.5	1	1	9.286557
## 1150	42 34.1	0	1	8.696131
## 1151	18 30.3	0	1	7.697911
## 1152	58 36.5	0	1	9.412125
## 1153	43 32.6	3	2	10.619894
## 1155	48 27.9	4	1	9.307029
## 1156	36 22.1	3	1	8.885748
## 1158	23 23.2	2	1	9.576792
## 1159	20 30.6	0	1	7.807803
## 1160	32 41.1	0	1	8.291506
## 1161	43 34.6	1	1	8.952508
## 1165	41 28.3	1	1	8.875364
## 1166	35 26.1	0	1	8.561782
## 1167	57 40.4	0	1	9.304058
## 1168	29 24.6	2	1	8.418362
## 1170	37 34.1	1	1	8.718067
## 1171	18 27.4	1	2	9.751424
## 1174	38 29.3	2	1	8.773050
## 1176	22 27.1	0	1	7.675249
## 1178	40 27.4	1	1	8.779079
## 1179	23 34.9	0	1	7.972290
## 1180	31 29.8	0	2	9.870467
## 1181	42 41.3	1	1	8.942562
## 1182	24 29.9	0	1	7.955313
## 1183	25 30.3	0	1	7.875875
## 1184	48 27.4	1	1	9.153493
## 1185	23 28.5	1	2	9.816198
## 1186	45 23.6	2	1	9.059962
## 1188	62 32.7	0	1	9.535665
## 1190	23 28.0	0	1	9.482402
## 1192	41 21.8	1	1	9.527009
## 1193	58 32.4	1	1	9.474177

## 1195	31 21.8	0	1	8.327020
## 1196	19 27.9	3	1	9.843669
## 1197	19 30.0	0	2	10.413539
## 1198	41 33.6	0	1	8.648193
## 1199	40 29.4	1	1	8.763053
## 1200	31 25.8	2	1	8.504049
## 1201	37 24.3	2	1	8.732103
## 1202	46 40.4	2	1	9.074891
## 1203	22 32.1	0	1	7.628187
## 1204	51 32.3	1	1	9.206740
## 1205	18 27.3	3	2	9.810465
## 1206	35 17.9	1	1	8.540226
## 1207	59 34.8	2	1	10.516254
## 1209	37 25.6	1	2	9.918221
## 1210	59 37.1	1	1	9.421182
## 1211	36 30.9	1	1	8.589209
## 1212	39 34.1	2	1	10.067434
## 1216	18 39.1	0	1	9.464212
## 1217	40 25.1	0	1	8.597050
## 1219	46 34.6	1	2	10.637335
## 1221	30 21.9	1	1	8.459183
## 1222	40 25.0	2	1	8.793841
## 1223	50 25.3	0	1	9.041054
## 1224	20 24.4	0	2	10.170674
## 1225	41 23.9	1	1	8.833241
## 1227	38 16.8	2	1	8.800949
## 1232	20 21.8	0	2	9.911820
## 1233	54 24.6	3	1	9.431859
## 1235	45 27.8	2	1	9.049674
## 1236	26 31.1	0	1	7.900848
## 1237	63 21.7	0	1	9.578439
## 1239	37 22.7	3	1	8.851593
## 1241	52 41.8	2	2	10.763628
## 1244	28 33.1	0	1	8.061995
## 1245	18 33.3	0	1	7.035216
## 1247	45 25.7	3	1	9.116227
## 1248	33 29.4	4	1	8.709328
## 1249	18 39.8	0	1	7.398762
## 1250	32 33.6	1	2	10.534960
## 1251	24 29.8	0	2	9.833517
## 1253	20 27.3	0	2	9.694792
## 1255	34 27.7	0	1	8.392799
## 1256	42 37.9	0	1	8.775551
## 1257	51 36.4	3	1	9.344586
## 1258	54 27.6	1	1	9.333083
## 1260	52 23.2	0	1	9.229924
## 1261	32 20.5	0	1	8.421614
## 1262	28 37.1	1	1	8.094732
## 1264	43 29.9	1	1	8.900788
## 1265	49 33.3	2	1	9.246760

## 1266	64 23.8	0	2 10.200867
## 1267	55 30.5	0	1 9.278417
## 1268	24 31.1	0	2 10.441560
## 1269	20 33.3	0	1 7.539288
## 1272	25 34.5	0	1 8.013611
## 1273	43 25.5	5	1 9.580408
## 1274	35 27.6	1	1 8.465279
## 1275	26 27.1	0	2 9.743515
## 1277	22 30.4	0	1 7.916425
## 1278	32 29.7	0	1 8.379548
## 1279	39 29.9	1	2 10.019582
## 1280	25 26.8	2	1 8.340244
## 1281	48 33.3	0	1 9.022043
## 1282	47 27.6	2	2 10.107884
## 1283	18 21.7	0	2 9.566858
## 1284	18 30.0	1	1 7.450283
## 1285	61 36.3	1	2 10.766459
## 1286	47 24.3	0	1 9.051892
## 1287	28 17.3	0	1 8.224868
## 1288	36 25.9	1	1 8.607482
## 1289	20 39.4	2	2 10.554368
## 1290	44 34.3	1	1 8.874514
## 1291	38 20.0	2	1 8.872613
## 1292	19 34.9	0	2 10.458196
## 1293	21 23.2	0	1 7.323395
## 1294	46 25.7	3	1 9.137973
## 1296	20 22.0	1	1 7.583136
## 1297	18 26.1	0	1 7.443623
## 1298	28 26.5	2	1 8.375731
## 1299	33 27.5	2	1 8.568166
## 1300	19 25.7	1	1 7.905010
## 1301	45 30.4	0	2 11.044407
## 1302	62 30.9	3	2 10.751888
## 1304	43 27.8	0	2 10.540850
## 1306	24 27.7	0	1 7.809793
## 1307	29 21.9	0	2 9.687524
## 1308	32 28.1	4	2 9.974527
## 1310	41 32.2	2	1 8.835787
## 1311	42 26.3	1	1 8.845188
## 1314	19 34.7	2	2 10.502258
## 1316	18 28.3	1	1 9.330106
## 1317	19 20.6	0	1 7.456847
## 1318	18 53.1	0	1 7.059154
## 1319	35 39.7	4	1 9.878002
## 1321	31 31.1	3	1 8.598777
## 1322	62 26.7	0	2 10.243572
## 1325	31 25.9	1	1 8.352293
## 1326	61 33.5	0	1 9.483670
## 1327	42 32.9	0	1 8.860786
## 1328	51 30.0	1	1 9.146111

## 1329	23	24.2	2	1	10.016626
## 1330	52	38.6	2	1	9.242344
## 1332	23	33.4	0	1	9.286925
## 1334	50	31.0	3	1	9.268661
## 1335	18	31.9	0	1	7.698927
## 1336	18	36.9	0	1	7.396231
## 1337	21	25.8	0	1	7.604870
## 1338	61	29.1	0	2	10.279914

Dados_Testes

##	idade	bmi	filhos	fumante	gastos
## 2	18	33.8	1	1	7.453301
## 3	28	33.0	3	1	8.400538
## 6	31	25.7	0	1	8.231275
## 9	37	29.8	2	1	8.765054
## 21	60	36.0	0	1	9.490155
## 22	30	32.4	1	1	8.330801
## 23	18	34.1	0	1	7.036157
## 25	37	28.0	2	1	8.732933
## 27	63	23.1	0	1	9.578577
## 28	55	32.8	2	1	9.414801
## 32	18	26.3	0	1	7.695390
## 42	31	36.6	2	1	8.507094
## 48	28	34.8	0	1	8.176650
## 49	60	24.5	0	1	9.443822
## 50	36	35.2	1	2	10.563832
## 56	58	37.0	2	2	10.768411
## 59	53	22.9	1	2	10.053836
## 61	43	27.4	3	1	9.060240
## 64	28	25.9	1	1	8.326914
## 68	40	26.3	1	1	8.762393
## 70	28	24.0	3	2	9.779235
## 71	27	24.8	0	2	9.715819
## 78	21	35.5	0	1	7.334636
## 79	22	39.8	0	1	7.921180
## 80	41	33.0	0	1	8.790424
## 81	31	26.9	1	1	8.398682
## 83	22	37.6	1	2	10.523127
## 86	45	22.9	2	2	9.956960
## 88	56	27.2	0	1	9.312281
## 90	55	27.0	0	1	9.313130
## 98	55	38.3	0	1	9.232716
## 99	56	20.0	0	2	10.017381
## 104	61	29.9	3	2	10.339876
## 108	26	30.9	2	1	8.262894
## 110	63	35.1	0	2	10.759084
## 112	55	29.7	2	1	9.382726
## 114	21	35.7	0	1	7.785193
## 115	52	32.2	3	1	9.349086

## 118	29 27.9	1	2 9.857851
## 121	44 37.1	2	1 8.954201
## 124	44 31.4	1	2 10.585485
## 130	38 34.7	2	1 8.713156
## 131	59 26.5	0	1 9.458406
## 136	22 28.1	0	1 7.675861
## 137	19 34.1	0	1 7.140009
## 138	22 25.2	0	1 7.623490
## 148	51 37.7	1	1 9.198026
## 153	32 37.1	3	1 8.753741
## 155	40 25.5	1	1 8.864632
## 156	44 39.5	0	1 8.846310
## 170	27 18.9	3	1 8.482167
## 175	24 33.3	0	1 7.956981
## 178	54 29.2	1	1 9.253026
## 181	58 28.6	0	1 9.370406
## 184	44 26.4	0	1 8.911864
## 186	36 41.9	3	2 10.686323
## 187	26 29.9	2	1 8.289534
## 191	61 31.6	0	1 9.438082
## 193	25 25.7	0	1 7.667462
## 195	18 34.4	0	1 7.036562
## 196	19 30.6	0	1 7.402183
## 197	39 32.8	0	1 8.639361
## 199	51 18.1	0	1 9.174117
## 200	64 39.3	0	1 9.609219
## 201	19 32.1	0	1 7.664196
## 204	27 36.1	0	2 10.522286
## 209	63 31.8	0	1 9.538273
## 215	45 30.9	2	1 9.050175
## 216	41 37.1	2	1 8.905413
## 218	27 23.1	0	1 7.817521
## 230	47 25.5	2	1 9.129701
## 234	59 27.5	1	1 9.420101
## 239	19 29.1	0	2 9.761502
## 240	44 38.1	1	1 8.875241
## 241	23 36.7	2	2 10.558716
## 245	63 27.7	0	2 10.292931
## 246	54 30.0	0	1 10.105468
## 248	24 35.9	0	1 7.594346
## 256	55 25.4	3	1 9.476339
## 260	19 31.9	0	2 10.426744
## 263	52 24.3	3	2 10.121411
## 265	53 38.1	3	1 9.926374
## 266	46 42.4	3	2 10.739677
## 269	45 30.2	1	1 8.914767
## 272	50 34.2	2	2 10.665621
## 277	19 20.6	2	1 7.938695
## 280	51 21.6	1	1 9.195747
## 283	30 27.6	1	1 8.351641

## 290	52 26.4	3	1	10.165576
## 291	28 33.4	0	1	8.062124
## 292	29 29.6	1	1	9.917282
## 294	22 28.8	0	1	7.676358
## 297	19 27.7	0	2	9.698788
## 300	48 28.9	1	1	9.132325
## 301	36 27.6	3	1	8.816815
## 304	28 33.0	2	1	8.377807
## 309	58 34.9	0	1	9.388034
## 310	41 33.1	2	1	8.955340
## 318	54 32.8	0	1	9.252928
## 322	26 29.6	4	1	10.113410
## 323	34 30.8	0	2	10.477052
## 324	57 40.9	0	1	9.355851
## 326	40 34.1	1	1	8.794857
## 330	52 36.7	0	1	9.120916
## 332	52 27.4	0	2	10.102077
## 340	46 27.7	1	1	9.015862
## 341	24 27.6	0	1	9.849835
## 342	62 30.0	0	1	9.499429
## 343	60 27.6	0	1	9.489266
## 347	33 35.8	2	1	8.494948
## 348	46 33.3	1	1	9.028154
## 349	36 29.9	1	1	8.608503
## 350	19 27.8	0	1	7.399844
## 351	57 23.2	0	1	9.378446
## 356	46 27.6	0	1	10.110626
## 358	47 29.8	3	1	9.171634
## 359	23 41.9	0	1	7.516041
## 362	35 30.5	1	1	8.466125
## 363	19 21.7	0	2	9.535644
## 374	26 32.9	2	2	10.493639
## 379	64 30.1	3	1	9.708428
## 384	35 43.3	2	1	8.673670
## 386	19 34.4	0	1	7.140342
## 391	48 35.6	4	1	9.281439
## 396	46 19.9	0	1	8.926213
## 397	43 34.4	3	1	9.050406
## 403	64 33.0	0	1	9.595104
## 407	33 24.3	0	1	8.339286
## 412	44 20.2	1	2	9.883020
## 414	25 23.9	5	1	8.533086
## 416	43 35.6	1	1	8.901874
## 421	64 33.9	0	2	10.755544
## 422	61 35.9	0	2	10.749337
## 426	45 24.3	5	1	9.189001
## 427	38 27.3	1	1	8.787994
## 433	42 26.9	0	1	8.694455
## 436	60 33.1	3	1	9.541069
## 437	22 31.7	0	1	7.720817

## 444	59 36.5	1	1 10.250189
## 445	56 26.7	1	2 10.170048
## 450	35 38.6	1	1 8.468492
## 458	57 30.5	0	1 9.379305
## 464	56 25.9	0	1 9.320577
## 466	30 28.4	1	2 9.879296
## 468	56 33.8	2	1 9.444889
## 470	18 24.1	1	1 7.696713
## 478	25 35.6	0	1 7.837708
## 485	48 34.3	3	1 9.165660
## 496	20 27.9	0	1 7.584275
## 497	31 23.6	2	1 8.503429
## 501	29 34.4	0	2 10.496751
## 503	51 23.2	1	2 10.008663
## 507	22 31.4	1	1 7.879772
## 511	56 32.1	1	1 9.372714
## 512	27 33.7	0	1 7.823410
## 515	39 28.3	1	2 9.956182
## 519	35 31.0	1	1 8.564224
## 521	50 27.4	0	1 10.152555
## 523	51 33.9	0	1 9.196880
## 532	62 31.7	0	1 9.549914
## 536	38 28.0	1	1 8.710641
## 537	33 38.9	3	1 8.694901
## 538	46 30.2	2	1 9.085354
## 548	54 46.7	2	1 9.353438
## 553	62 21.4	0	1 9.469401
## 555	25 41.3	0	1 9.791377
## 560	19 35.5	0	1 7.406365
## 561	46 20.0	2	1 9.126289
## 571	31 29.1	0	1 8.232517
## 573	30 43.1	2	1 8.466666
## 581	59 25.5	1	1 9.466066
## 584	32 23.7	1	1 9.777144
## 587	21 20.2	3	1 8.258736
## 593	20 31.1	2	1 7.850287
## 596	46 33.7	1	1 9.085229
## 597	42 29.5	2	1 8.941193
## 606	51 34.1	0	1 9.136000
## 618	49 25.6	2	2 10.056490
## 619	19 33.1	0	2 10.446970
## 625	59 28.8	0	1 9.403405
## 630	44 39.0	0	2 10.668571
## 631	53 36.1	1	1 9.218889
## 634	40 22.7	2	1 8.878129
## 637	19 24.5	1	1 7.904375
## 640	56 33.7	4	1 9.468786
## 646	48 30.8	3	1 9.224356
## 648	40 23.4	3	1 9.018245
## 649	18 28.5	0	1 7.445552

## 655	59 35.2	0	1 9.412835
## 656	52 25.3	2	2 10.113239
## 659	48 35.9	1	1 10.180826
## 661	37 46.5	3	1 8.769603
## 662	57 24.0	1	1 10.007507
## 667	49 28.7	1	1 9.071476
## 672	29 31.2	0	1 8.279849
## 673	36 29.7	0	1 8.389298
## 675	44 43.9	2	2 10.740757
## 677	55 40.8	3	1 9.432347
## 681	21 17.4	1	1 7.857585
## 683	39 35.3	2	2 10.599229
## 684	53 24.3	0	1 9.196593
## 685	33 18.5	1	1 8.469267
## 687	42 26.1	2	1 8.952819
## 688	40 41.7	0	1 8.601305
## 689	47 24.1	1	1 10.174910
## 694	24 23.7	0	1 7.763434
## 695	27 34.8	1	1 8.182559
## 701	21 34.9	0	1 7.611125
## 704	34 26.4	1	1 8.591436
## 710	36 27.7	0	1 8.606853
## 717	49 22.6	1	1 9.166074
## 726	30 39.1	3	2 10.619678
## 730	41 36.1	1	1 8.821931
## 733	24 30.1	3	1 8.351122
## 734	48 27.3	1	1 9.153479
## 738	26 23.7	2	1 8.156031
## 742	27 29.2	0	2 9.811729
## 747	34 27.0	2	1 9.370574
## 750	28 30.9	0	1 8.026990
## 752	21 29.0	0	1 7.552951
## 757	39 22.8	3	1 8.985423
## 765	45 25.2	2	1 9.115488
## 766	57 31.8	0	1 9.379460
## 769	64 39.7	0	1 9.569345
## 770	38 19.5	2	1 8.844083
## 772	53 26.7	2	1 9.319265
## 774	19 28.9	0	2 9.784057
## 780	53 28.9	0	1 9.197236
## 782	18 41.1	0	1 7.044731
## 784	50 27.6	1	2 10.107255
## 792	19 27.6	0	1 7.132825
## 805	23 26.5	0	1 7.504325
## 806	45 35.8	0	1 8.953105
## 808	19 36.6	0	1 7.667102
## 812	33 42.9	3	1 8.757939
## 815	36 34.4	2	1 8.627716
## 816	20 31.5	0	1 7.537925
## 819	47 26.1	1	2 10.060547

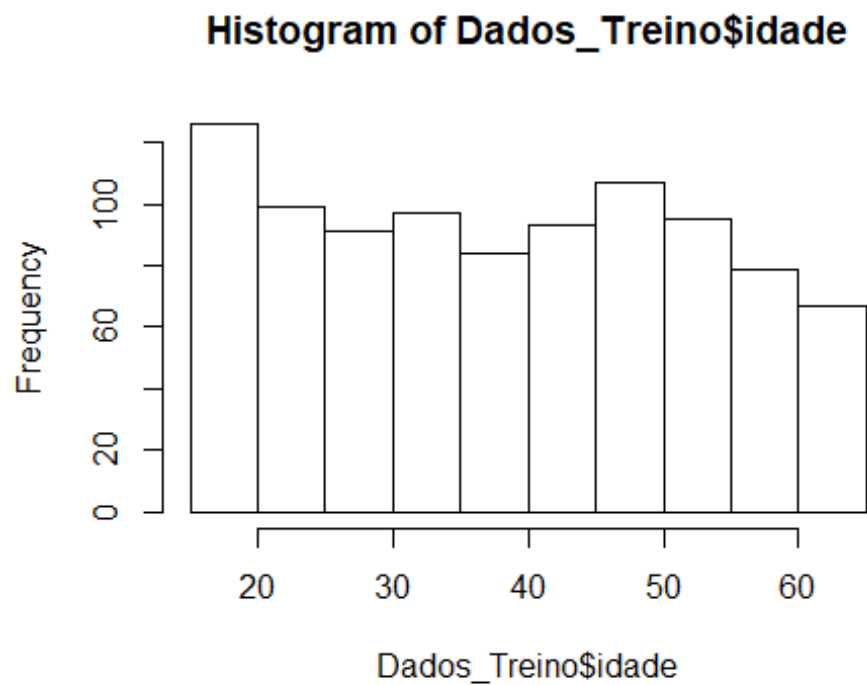
## 820	33 35.5	0	2	10.917547
## 821	45 33.7	1	1	8.915422
## 825	60 24.3	0	1	9.435370
## 832	36 25.8	0	1	8.569097
## 835	36 33.8	1	1	8.589971
## 838	56 28.3	0	1	9.363724
## 839	35 23.5	2	1	8.764411
## 840	59 31.4	0	1	9.443211
## 842	59 24.7	0	1	9.419299
## 845	53 30.5	0	1	9.217521
## 846	60 32.5	0	2	10.714617
## 850	55 32.8	0	1	9.268763
## 851	37 30.8	0	2	10.525948
## 852	61 32.3	2	1	9.555321
## 853	46 35.5	0	2	10.648080
## 854	53 23.8	2	1	9.369878
## 861	37 47.6	2	2	10.738861
## 862	38 28.0	3	1	8.875020
## 867	18 37.3	0	1	7.040055
## 870	25 24.3	3	1	8.387460
## 875	44 21.9	3	1	9.092811
## 879	41 28.8	1	1	8.745482
## 889	22 39.5	0	1	7.428095
## 890	57 33.6	1	1	9.388079
## 892	36 29.0	4	1	8.887903
## 895	62 32.1	0	1	9.514511
## 897	43 20.0	2	2	9.893339
## 900	19 22.5	0	1	7.657916
## 902	60 40.9	0	2	10.792891
## 905	60 35.1	0	1	9.444985
## 907	27 32.6	3	1	8.486099
## 912	18 31.7	0	2	10.426223
## 919	61 28.2	0	1	9.475924
## 929	62 39.2	0	1	9.508280
## 930	41 34.2	1	1	8.746677
## 934	45 35.3	0	1	8.902202
## 935	32 37.2	2	1	8.449640
## 938	39 24.2	5	1	9.101173
## 939	18 26.2	2	1	7.742402
## 940	53 29.5	0	1	9.157745
## 942	50 46.1	1	1	9.164251
## 943	18 40.2	0	1	7.704122
## 946	56 35.8	1	1	9.365131
## 949	42 31.3	0	1	8.757592
## 951	57 18.3	0	1	9.353130
## 957	54 30.8	1	2	10.645413
## 964	46 24.8	3	1	9.159107
## 968	44 25.4	1	1	8.925059
## 976	29 22.9	0	2	9.688979
## 977	48 40.2	0	1	8.962412

## 984	27 30.6	1	1	9.728920
## 986	44 25.8	1	1	8.939139
## 988	45 27.6	1	1	10.252036
## 992	38 27.8	2	1	8.874148
## 994	38 28.3	1	1	8.609676
## 997	39 34.1	3	1	8.911735
## 1001	30 23.0	2	2	9.762026
## 1003	24 25.8	0	1	7.587285
## 1007	28 29.3	2	1	8.398018
## 1009	25 25.0	2	1	10.053693
## 1010	51 27.7	1	1	9.206103
## 1011	48 22.8	0	1	9.020274
## 1013	61 33.3	4	1	10.507265
## 1014	48 32.3	1	1	9.078550
## 1015	38 27.6	0	1	8.591101
## 1023	47 36.1	1	2	10.650439
## 1024	18 23.3	1	1	7.444851
## 1031	46 23.7	1	2	9.984020
## 1033	30 27.9	0	1	8.327852
## 1036	54 23.0	3	1	9.400504
## 1037	22 37.1	2	2	10.531681
## 1039	22 28.9	0	1	7.719059
## 1042	18 23.1	0	1	7.441144
## 1044	28 25.8	0	1	8.058786
## 1046	43 24.7	2	2	9.993366
## 1050	49 30.9	0	2	10.589802
## 1055	27 21.5	0	1	8.117751
## 1058	45 31.8	0	1	9.794192
## 1059	24 39.5	0	1	7.816409
## 1061	24 32.0	0	1	7.591650
## 1069	63 21.7	1	1	9.571495
## 1070	54 31.9	1	1	9.299161
## 1072	63 31.4	0	1	9.544987
## 1073	21 31.3	0	1	7.554612
## 1074	54 28.9	2	1	9.400684
## 1076	32 29.6	1	1	8.425701
## 1079	28 31.7	0	2	10.453692
## 1083	38 20.0	1	1	8.675205
## 1086	39 18.3	5	2	9.853418
## 1088	57 31.5	0	1	9.337258
## 1093	23 32.7	3	1	8.186320
## 1094	22 30.4	0	2	10.431393
## 1096	18 31.4	4	1	8.425339
## 1097	51 35.0	2	2	10.706412
## 1099	52 30.9	0	1	10.045229
## 1103	29 38.9	1	1	8.152316
## 1104	58 36.1	0	1	9.338142
## 1105	37 29.8	0	1	9.924299
## 1108	50 26.2	2	1	9.258554
## 1109	26 30.0	1	1	7.973875

## 1112	38 38.4	3	2 10.644216
## 1113	48 25.9	3	2 10.093320
## 1115	23 24.5	0	1 7.781598
## 1117	41 29.6	5	1 9.129391
## 1119	33 35.8	1	2 10.552755
## 1128	35 35.9	2	1 8.671890
## 1134	52 18.3	0	1 9.209444
## 1138	26 22.2	0	1 8.063469
## 1139	33 30.3	0	1 8.217263
## 1141	50 37.1	1	1 9.110302
## 1147	60 32.8	0	2 10.870297
## 1148	20 31.9	0	1 7.723815
## 1154	35 35.8	1	1 8.635946
## 1157	19 44.9	0	2 10.589679
## 1162	34 42.1	2	1 8.541728
## 1163	30 38.8	1	1 9.850254
## 1164	18 28.2	0	1 7.696590
## 1169	32 35.2	2	1 8.449051
## 1172	43 26.7	2	2 10.020319
## 1173	56 41.9	0	1 9.314125
## 1175	29 32.1	2	1 8.397039
## 1177	52 24.1	1	2 10.081117
## 1187	20 35.6	3	2 10.531172
## 1189	43 25.3	1	2 9.988350
## 1191	31 32.8	2	1 8.580619
## 1194	48 36.6	0	1 9.067761
## 1208	36 33.4	2	2 10.556216
## 1213	18 21.5	0	1 7.439830
## 1214	52 33.3	2	1 9.287935
## 1215	27 31.3	1	1 8.283006
## 1218	29 37.3	2	1 8.308475
## 1220	38 30.2	3	1 8.927601
## 1226	33 39.8	1	1 8.475467
## 1228	42 37.2	2	1 8.876546
## 1229	56 34.4	0	1 9.268065
## 1230	58 30.3	0	1 9.387504
## 1231	52 34.5	3	2 11.002456
## 1234	58 23.3	0	1 9.336578
## 1238	58 28.2	0	1 9.411185
## 1240	25 42.1	1	1 8.082847
## 1242	64 37.0	2	2 10.811296
## 1243	22 21.3	3	1 8.365502
## 1246	28 24.3	5	1 8.633263
## 1252	19 19.8	0	1 7.124132
## 1254	40 29.3	4	1 9.669588
## 1259	55 37.7	3	1 10.311070
## 1263	41 28.1	1	1 8.820284
## 1270	45 27.5	3	1 9.061295
## 1271	26 33.9	1	1 8.099412
## 1276	57 23.7	0	1 9.301946

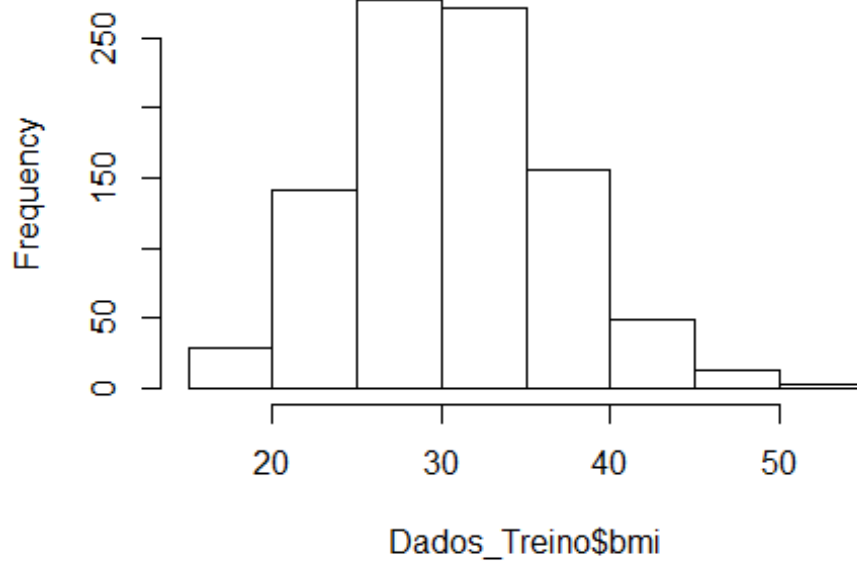
```
## 1295    58 25.2    0    1  9.386906
## 1303    25 20.8    1    1  8.073649
## 1305    42 24.6    2    2  9.964553
## 1309    25 30.2    0    2 10.431189
## 1312    33 26.7    0    1  8.427577
## 1313    34 42.9    1    1  8.419858
## 1315    30 23.7    3    2  9.839796
## 1320    39 26.3    2    1  8.882072
## 1323    62 38.8    0    1  9.471269
## 1324    42 40.4    2    2 10.689587
## 1331    57 25.7    2    1  9.443764
## 1333    52 44.7    3    1  9.342394
```

```
hist(Dados_Treino$idade)
```



```
hist(Dados_Treino$bmi)
```

Histogram of Dados_Treino\$bmi



```
round(prop.table(table(Dados_Treino$filhos)) * 100, digits = 1)

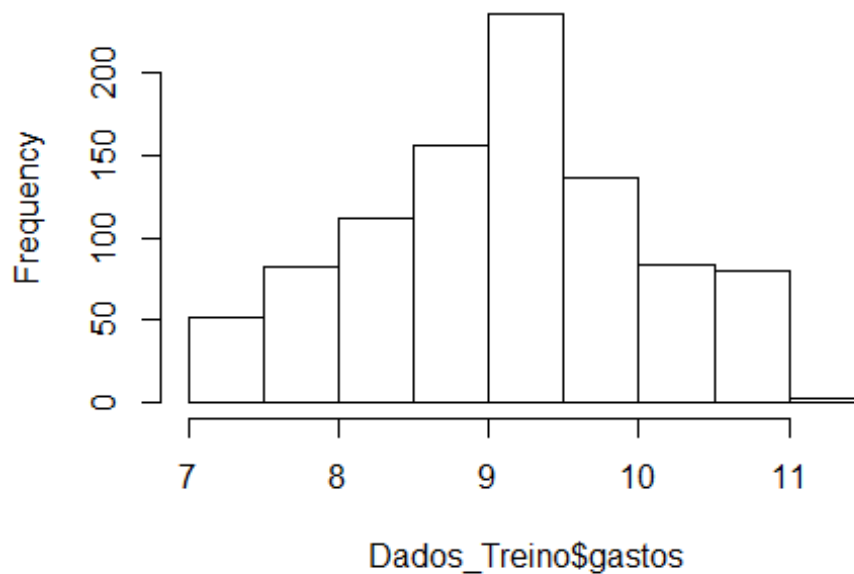
##
##      0      1      2      3      4      5
## 42.8 24.5 17.8 11.7  1.9  1.3

round(prop.table(table(Dados_Treino$fumante)) * 100, digits = 1)

##
##      1      2
## 79.4 20.6

hist(Dados_Treino$gastos)
```


Histogram of Dados_Treino\$gastos



```
# Fazemos as predições nos dados de teste
# O modelo utilizado aqui é o de Regressão Linear. No modelo1 estou consid
erando
# todas as variáveis do dataset.
Modelo_v3 <- train(gastos ~ ., data = Dados_Treino, method = "lm")
```

Etapa 4 -Interpretando o Modelo

Nesta etapa iremos analisar o resumo dos parâmetros do Modelo_v3.

```
summary(Modelo_v3)

##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15461 -0.20140 -0.05480  0.07138  2.04405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.471365   0.094988  57.601  < 2e-16 ***
## idade       0.033729   0.001075  31.377  < 2e-16 ***
## bmi         0.011047   0.002464   4.483 8.27e-06 ***
```

```
## filhos      0.101445    0.012507    8.111 1.57e-15 ***
## fumante     1.548701    0.037130   41.711 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4594 on 933 degrees of freedom
## Multiple R-squared:  0.7505, Adjusted R-squared:  0.7494
## F-statistic: 701.5 on 4 and 933 DF,  p-value: < 2.2e-16
```

Avaliando os parâmetros:

O primeiro parâmetro a ser analisado é o nível de significância de acordo com o p-value. O nível de significância é o limite para o p-valor, abaixo do qual assume-se que a hipótese nula é falsa. O p-valor é a probabilidade de se obter uma estatística de teste igual ou mais extrema que a estatística observada a partir de uma amostra de uma população quando ela é verdadeira. Isto significa que o nível de significância é a probabilidade de se rejeitar incorretamente a hipótese nula quando ela é verdadeira. O nível de significância corresponde ao erro do tipo I, cujos valores mais comuns são 10%, 5% e 1%. Iremos considerar o de 5%. Pode-se observar que as variáveis idade, bmi, filhos, fumante, região têm o nível de significância < 10%. Já as outras variáveis não influenciam tanto no modelo. Nesse sentido teremos que elaborar outro modelo, retirando as variáveis que não influenciam ou pouco influenciam na variável dependente. Retirando essas variáveis os valores das variáveis significativas e da interceptação serão alterados.

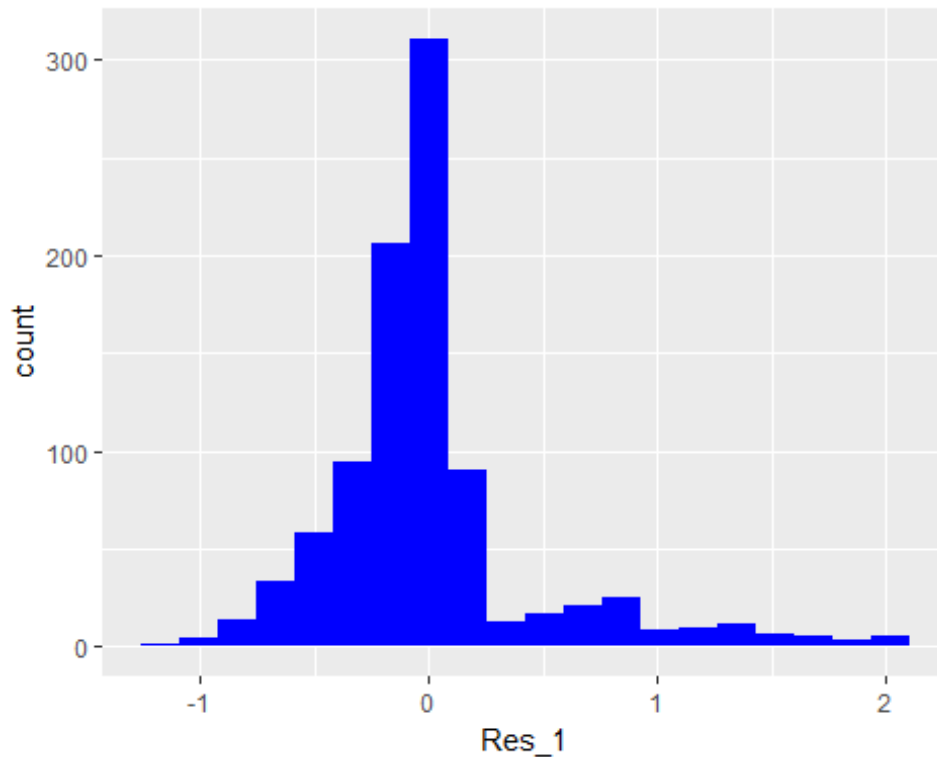
Comparando os parâmetros conforme pressuposições abaixo: Residuais

```
Res_1 <- residuals(Modelo_v3)

# Convertendo o objeto para um dataframe
Res_1 <- as.data.frame(Res_1)
head(Res_1)

##           Res_1
## X1    0.21635691
## X4    1.61421318
## X5   -0.15844136
## X7   -0.02517063
## X8    0.01473214
## X10   0.94359601

ggplot(Res_1, aes(Res_1)) +
  geom_histogram(bins = 20, fill = 'blue')
```



O histograma apresenta uma distribuição parecida com a normal, o que indica que a média entre os valores previstos e os valores observados é próximo de 0 (o que é bom).

Interpretação dos parâmetros:

Coefficients: Como pode ser observado os valores estimados mudaram no Modelo_v3. Com isso temos a equação da regressão linear múltipla: $\text{gastos} = 5.471365 + 0.033729 \times \text{idade} + 0.011047 \times \text{bmi} + 0.101445 \times \text{filhos} + 1.548701 \times \text{fumante}$

Exemplo: Linha 59 do dataset Despesas

```
Dados_Norm_3[59,]
##      idade  bmi filhos fumante  gastos
## 59     53 22.9      1        2 10.05384

gastos_59 = 5.471365 + 0.033729*53 + 0.011047*22.9 + 0.101445*1 + 1.548701*2
gastos_59
## [1] 10.71083
```

Residual standard error: Temos um $dp = 0.4595$ dos resíduos, e um grau de liberdade = 933

Multiple R-squared: Quanto mais próximo de 1 melhor. No nosso caso temos um $R^2 = 0.7505$ o que indica que 75,05% da variável dependente consegue ser explicada pelas variáveis explanatórias presentes no modelo.

Adjusted R-squared: Ele deve ser menor que R^2 . No nosso caso temos 0.7494.

F-statistic: Esse teste obtém os parâmetros do nosso modelo e compara com um modelo que tenha menos parâmetros

p-value: Temos uma probabilidade $< 2.2e-16$ que a variável não seja relevante.

Etapa 5 - Previsão

Prevendo despesas médicas Usando a função predict conseguimos realizar a previsão do modelo baseado nas no arquivo de teste

```
Previsao_v3 <- predict(Modelo_v3, Dados_Testes)
# Abaixo temos o cabeçalho das 10 primeiras previsões
head(Previsao_v3, 10)

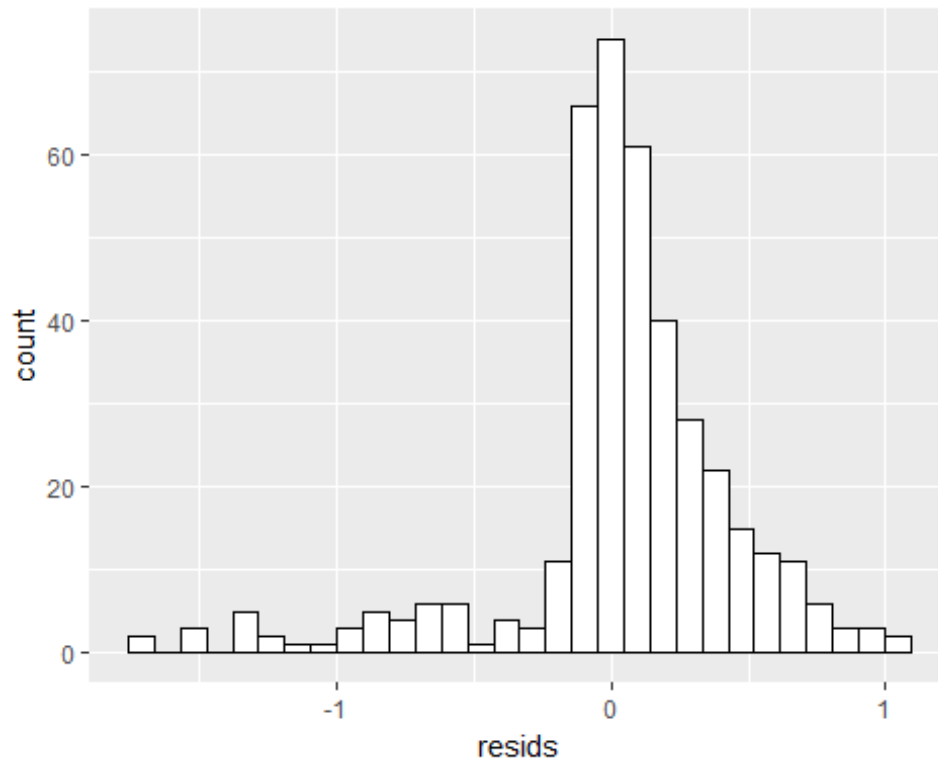
##          2          3          6          9         21         22         23         2
5
## 8.102010 8.633350 8.349561 8.800115 9.441478 8.491290 8.003879 8.78023
1
##          27         28
## 9.400161 9.440374

# Visualizando os valores previstos e observados
Resultados <- cbind(Dados_Testes$gastos, Previsao_v3)
colnames(Resultados) <- c('Real', 'Previsto')
Resultados <- as.data.frame(Resultados)
# Estamos prevendo os gastos e eles não podem ser negativos. Vamos verifi
car
min(Resultados)

## [1] 7.036157

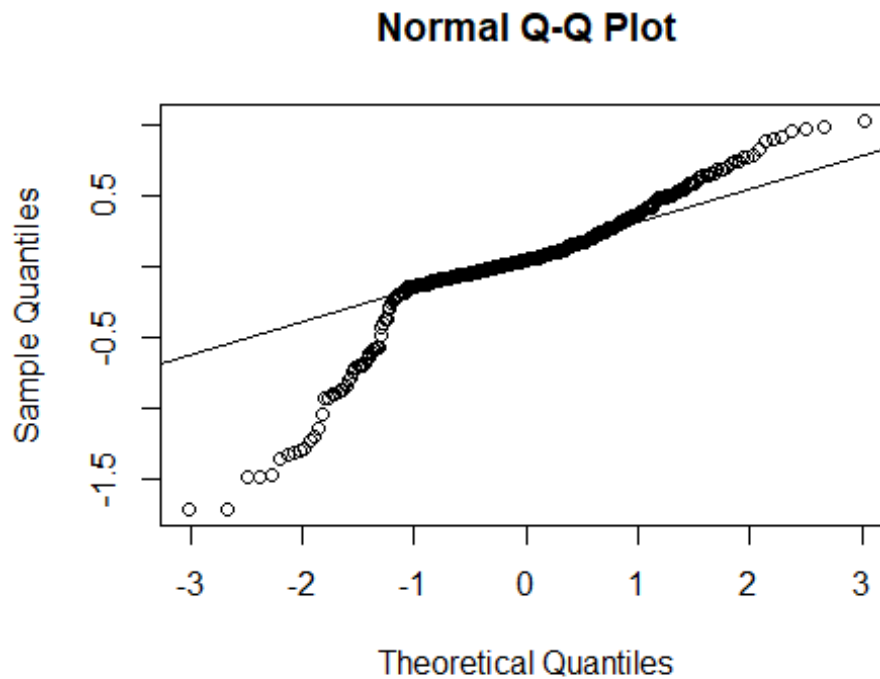
# Grafico dos residuos
Resultados_2 <- mutate(Resultados, resid = Previsto - Real)
ggplot(Resultados_2, aes(x = resid)) +
  geom_histogram(fill = "white", color = "black")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



O gráfico acima parece com uma normal porem a presença de outliers faz com que ele # fique mais enviesado para a esquerda. Mais abaixo irei realizar alguns testes afim de # amenizar esses erros

```
qqnorm(Resultados_2$resids)
qqline(Resultados_2$resids)
```



Etapa 6 - Avaliando a Performance

Calculando o erro médio: Quão distantes seus valores previstos estão dos valores observados. Serve para avaliar as versões do modelo

```
# MSE
mse <- mean((Resultados$Real - Resultados$Previsto)^2)
mse

## [1] 0.1811507

# RMSE
rmse <- mse^0.5
rmse

## [1] 0.4256181

# Calculando R Squared
SSE = sum((Resultados$Previsto - Resultados$Real)^2)
SST = sum((mean(Dados_Norm$gastos) - Resultados$Real)^2)

# R-Squared
# Ajuda a avaliar o nível de precisão do nosso modelo. Quanto maior, melhor, sendo 1 o valor ideal.
```

```
R2 = 1 - (SSE/SST)
R2
## [1] 0.7877985
```

Conclusão

Baseado na comparação realizada na versão anterior os Modelos v1 e v2 são muito parecidos. Porém o Modelo_v2 é um pouco melhor pois possui menos variáveis e tem o R ajustado um pouco maior. Neste modelo de machine learning podemos dizer que 75,04% da variável dependente consegue ser explicada pelo modelo. Nesta versão 3, conseguimos melhorar a performance do modelo por meio da normalização da variável gastos. Realizamos mais alguns testes mas o melhor do um R-squared de 0.7877, ou seja Neste modelo de machine learning (Modelo_v3) podemos dizer que 78,77% da variável dependente consegue ser explicada pelo modelo.

Evanil Tiengo Junior

Bello Horizonte, 06 de novembro de 2018