



Bioestatística

Atividade no R

Prof. Leticia Raposo

2019

Abalone

Haliotis é um gênero de moluscos gastrópodes marinhos da família *Haliotidae* e o único gênero catalogado desta família. Foi proposto por Linnaeus em 1758 e contém diversas espécies em águas costeiras de quase todo o mundo. Na gastronomia, o abalone é um molusco valorizado em países asiáticos. Wikipédia

O banco de dados disponível no arquivo *ExercicioAbalone.txt* contém 4177 observações e 9 variáveis. As descrições de cada variável são:

- **sexo:** M, F, e I (infantil)
- **comprimento:** comprimento (maior medida da concha), em mm
- **diametro:** diâmetro (perpendicular ao comprimento), em mm
- **altura:** altura, em mm
- **peso_total:** peso total, em gramas
- **peso_sem_concha:** peso sem a concha, em gramas
- **peso_intestinal:** peso intestinal (após sangramento), em gramas
- **peso_concha:** peso da concha depois de seca, em gramas
- **aneis:** número de anéis (1: até 10 anéis, 2: mais de 10 anéis)

Trabalhos mal formatados,
com caracteres estranhos e
não enviado nos formatos
Word ou HTML perdem 0,5.

```
dados <-  
read.table("C:/Users/Leticia/Google Drive/UNIRIO/Disciplinas Ministradas/2019.2/Ciências Ambientais -  
header = T)
```

1. Antes de iniciar a análise, verifique e responda:

- i. (0,5 ponto) *As variáveis foram lidas (codificadas) corretamente pelo R? Se não, faça a correta codificação.*

```
str(dados)
```

```
## 'data.frame':    4177 obs. of  9 variables:  
## $ sexo          : Factor w/ 3 levels "F","I","M": 2 2 2 2 2 3 3 2 2 2 ...  
## $ comprimento   : num  0.075 0.15 0.13 0.11 0.165 0.21 0.155 0.16 0.24 0.195 ...  
## $ diametro      : num  0.055 0.1 0.1 0.09 0.12 0.15 0.11 0.11 0.175 0.15 ...  
## $ altura        : num  0.01 0.025 0.03 0.03 0.03 0.05 0.04 0.025 0.065 0.045 ...  
## $ peso_total     : num  0.002 0.015 0.013 0.008 0.0215 0.0385 0.0155 0.018 0.0665 0.0375 ...  
## $ peso_sem_concha: num  0.001 0.0045 0.0045 0.0025 0.007 0.0155 0.0065 0.0065 0.031 0.018 ...  
## $ peso_intestinal: num  0.0005 0.004 0.003 0.002 0.005 0.0085 0.003 0.0055 0.0135 0.006 ...  
## $ peso_concha    : num  0.0015 0.005 0.004 0.003 0.005 0.01 0.005 0.005 0.017 0.011 ...  
## $ aneis         : int   1 1 1 1 1 1 1 1 1 1 ...
```

```
dados$aneis <- as.factor(dados$aneis)
```

Não, a variável aneis deveria ser qualitativa.

- ii. (1,0 ponto) *Há dados ausentes? Se sim, em qual(is) variável(is)?*

```
summary(dados)
```

O aluno precisa mostrar que
verificou as variáveis pelo
str. Se não tiver isso, retire
0,1.



```
## sexo      comprimento      diametro      altura
## F:1307    Min.      :0.075    Min.      :0.0550    Min.      :0.0000
## I:1342    1st Qu.:0.450    1st Qu.:0.3500    1st Qu.:0.1150
## M:1528    Median :0.545    Median :0.4250    Median :0.1400
##          Mean      :0.524    Mean      :0.4079    Mean      :0.1395
##          3rd Qu.:0.615    3rd Qu.:0.4800    3rd Qu.:0.1650
##          Max.      :0.815    Max.      :0.6500    Max.      :1.1300
## peso_total peso_sem_concha peso_intestinal peso_concha
## Min.      : 0.0020    Min.      : 0.001    Min.      :0.0005    Min.      :0.0015
## 1st Qu.: 0.4415    1st Qu.: 0.186    1st Qu.:0.0935    1st Qu.:0.1300
## Median : 0.7995    Median : 0.336    Median :0.1710    Median :0.2340
## Mean      : 2.8581    Mean      : 0.425    Mean      :0.1806    Mean      :0.2388
## 3rd Qu.: 1.3120    3rd Qu.: 0.502    3rd Qu.:0.2530    3rd Qu.:0.3290
## Max.      :28.2550    Max.      :13.485    Max.      :0.7600    Max.      :1.0050
## aneis
## 1:2096
## 2:2081
##
##
##
##
```

O aluno precisa mostrar que verificou a presença ou não de dados ausentes por algum comando. Se não fez isso, retire 0,1.

Não há dados ausentes.

2. Para a variável *comprimento*, pede-se:

- (0,6 ponto) Calcule a média aritmética, a mediana e a moda;
- (0,6 ponto) Calcule o primeiro e o terceiro quartis e também o IQR;
- (0,8 ponto) Calcule as medidas de dispersão (amplitude, variância, desvio-padrão e coeficiente de variação);

```
library(summarytools)
```

```
## Warning: package 'summarytools' was built under R version 3.6.1
```

```
## Registered S3 method overwritten by 'pryr':
```

```
## method      from
```

```
## print.bytes Rcpp
```

```
descr(dados$comprimento)
```

```
## Descriptive Statistics
```

```
## dados$comprimento
```

```
## N: 4177
```

```
##
```

```
## ----- comprimento -----
```

```
##
```

```
##          Mean      0.52
```

```
##          Std.Dev    0.12
```

```
##          Min        0.08
```

```
##          Q1         0.45
```

```
##          Median     0.55
```

```
##          Q3         0.62
```

```
##          Max        0.82
```

```
##          MAD        0.12
```

```
##          IQR        0.16
```

```
##          CV         0.23
```

Para as letras i, ii, iii, aluno pode calcular por qualquer forma, desde que seja por meio do R. Se ele não especificou cada valor, retire 0,1 em cada alternativa.



```
##           Skewness           -0.64
##          SE.Skewness           0.04
##           Kurtosis           0.06
##          N.Valid           4177.00
##          Pct.Valid           100.00
```

```
var(dados$comprimento) #variância
```

```
## [1] 0.01442231
```

```
diff(range(dados$comprimento)) #amplitude
```

```
## [1] 0.74
```

```
library(DescTools)
```

```
## Warning: package 'DescTools' was built under R version 3.6.1
```

```
Mode(dados$comprimento)
```

```
## [1] 0.550 0.625
```

- Média aritmética: 0,52 mm
- Mediana: 0,55 mm
- Moda: 0,550 e 0,625 mm
- Q1: 0,45 mm
- Q3: 0,62 mm
- IQR: 0,16 mm
- Amplitude: 0,74 mm
- Variância: 0,01 mm²
- Desvio-padrão: 0,12 mm

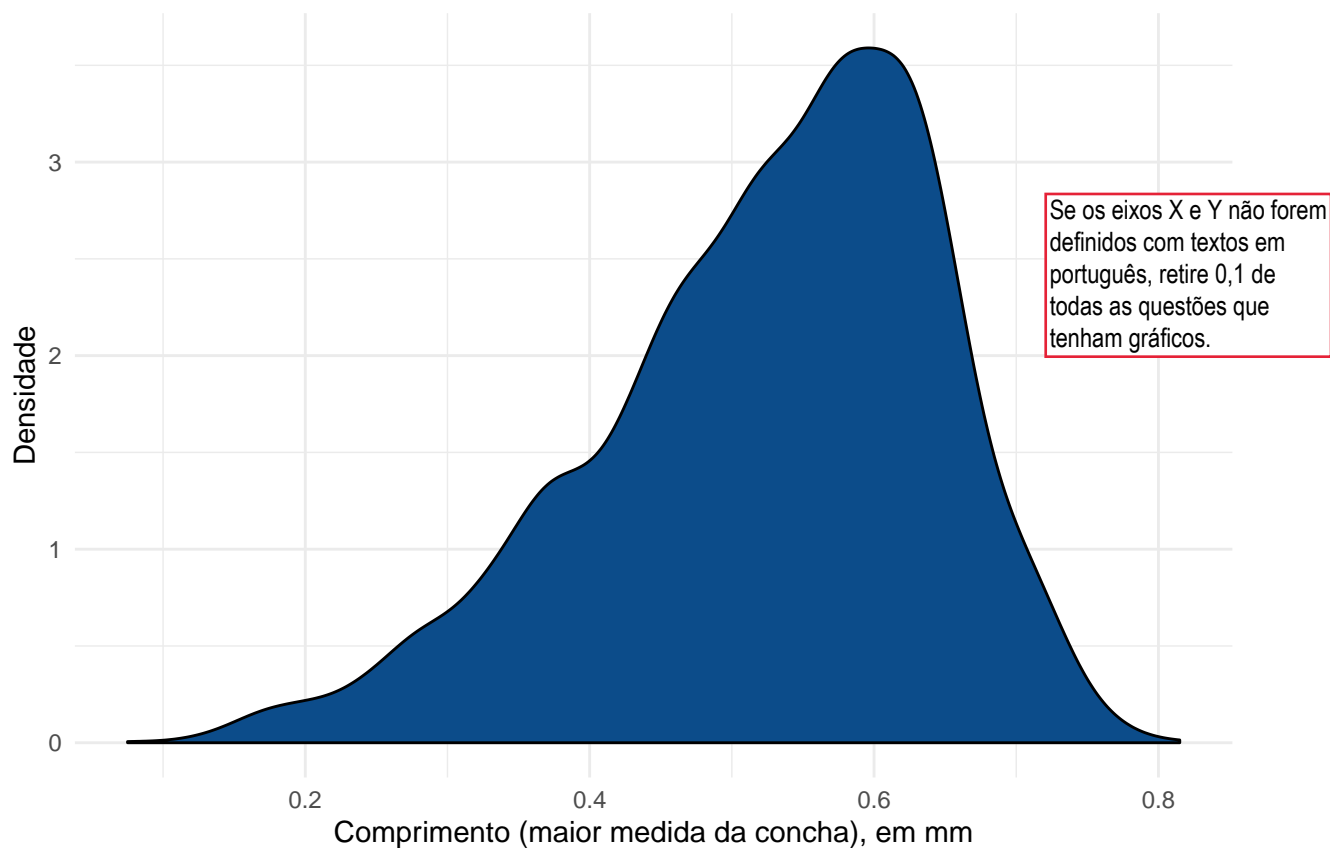
iv. (0,5 ponto) *Verifique se a distribuição é simétrica, assimétrica positiva ou assimétrica negativa;*

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
ggplot(dados) +
  aes(x = comprimento) +
  geom_density(adjust = 1L, fill = "#0c4c8a") +
  labs(x = "Comprimento (maior medida da concha), em mm", y = "Densidade") +
  theme_minimal()
```

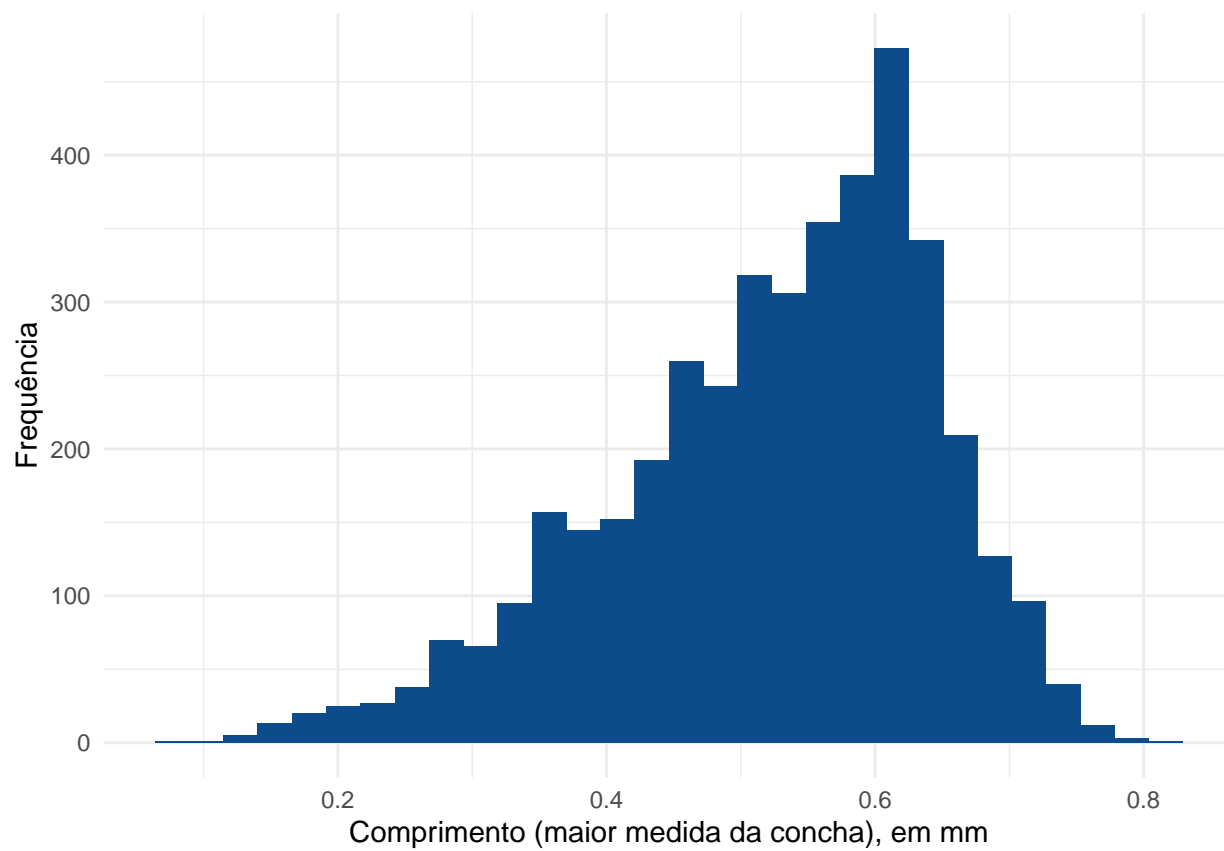
O aluno pode verificar por qualquer gráfico (histograma, densidade ou boxplot) ou medida de assimetria. Se não colocar gráfico ou medida, retire 0,1 e se não falar o tipo de assimetria, retire 0,1.



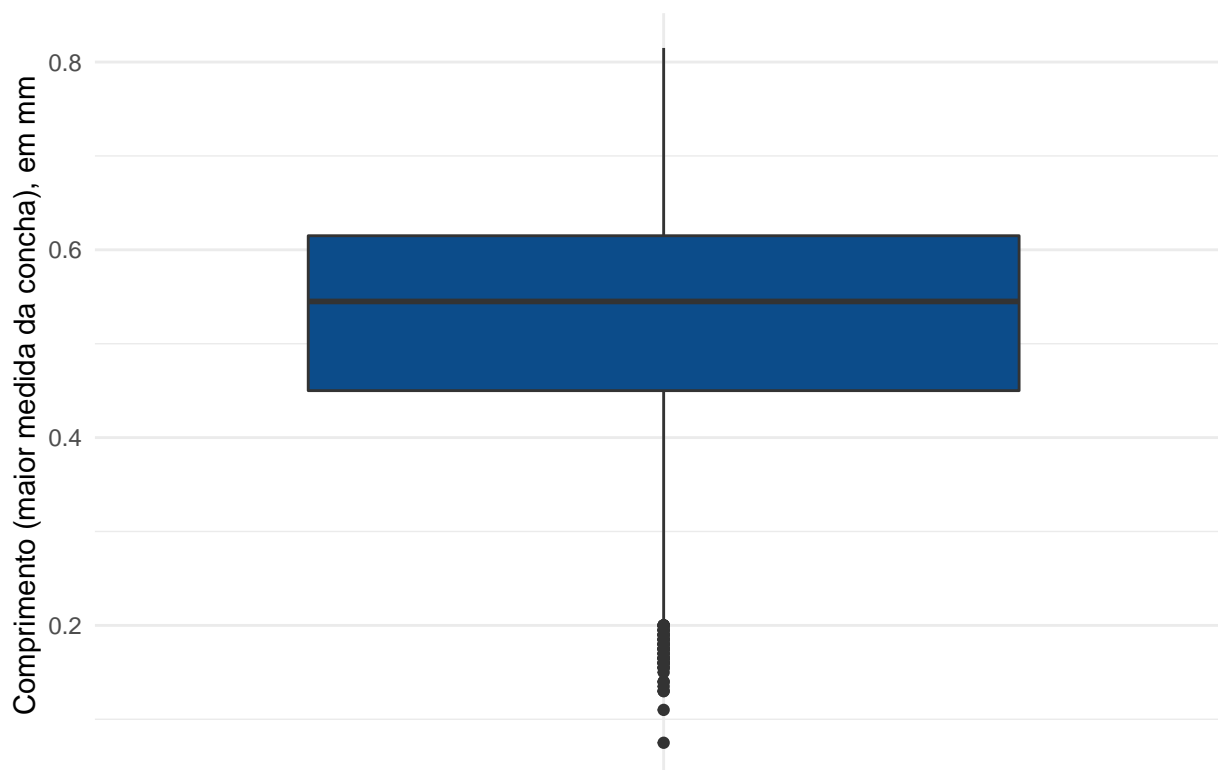
A distribuição é levemente assimétrica à esquerda.

v. (0,6 ponto) Construa o histograma e o boxplot para a variável em estudo;

```
library(ggplot2)
# Histograma
ggplot(dados) +
  aes(x = comprimento) +
  geom_histogram(bins = 30L, fill = "#0c4c8a") +
  labs(x = "Comprimento (maior medida da concha), em mm", y = "Frequência") +
  theme_minimal()
```



```
# Boxplot
ggplot(dados) +
  aes(x = "", y = comprimento) +
  geom_boxplot(fill = "#0c4c8a") +
  labs(x = " ", y = "Comprimento (maior medida da concha), em mm") +
  theme_minimal()
```



- vi. (0,4 ponto) *É possível observar outliers para a variável analisada?* Sim, há a presença de diversos outliers inferiores.

Basta uma resposta sim e já é o suficiente para ganhar a questão toda.

3. Para a variável *sexo*, pede-se:

- i. (0,7 ponto) *Construa uma tabela de distribuição de frequências. Comente o resultado.*

```
freq(dados$sexo)
```

```
## Frequencies
## dados$sexo
## Type: Factor
##
##      Freq  % Valid  % Valid Cum.  % Total  % Total Cum.
## -----
##      F    1307    31.29      31.29    31.29    31.29
##      I    1342    32.13      63.42    32.13    63.42
##      M    1528    36.58     100.00    36.58   100.00
##      <NA>      0      0.00      0.00     0.00   100.00
##      Total   4177   100.00     100.00   100.00   100.00
```

Pode ser qualquer tipo de tabela de distribuição de frequências com frequência absoluta e relativa. Se não tiver um comentário sobre a tabela, retire 0,4. Se não tiver a frequência relativa, retire 0,1.

Podemos observar que as frequências são próximas entre as 3 classes, apresentando um maior valor a classe adulta masculina (36,58%).

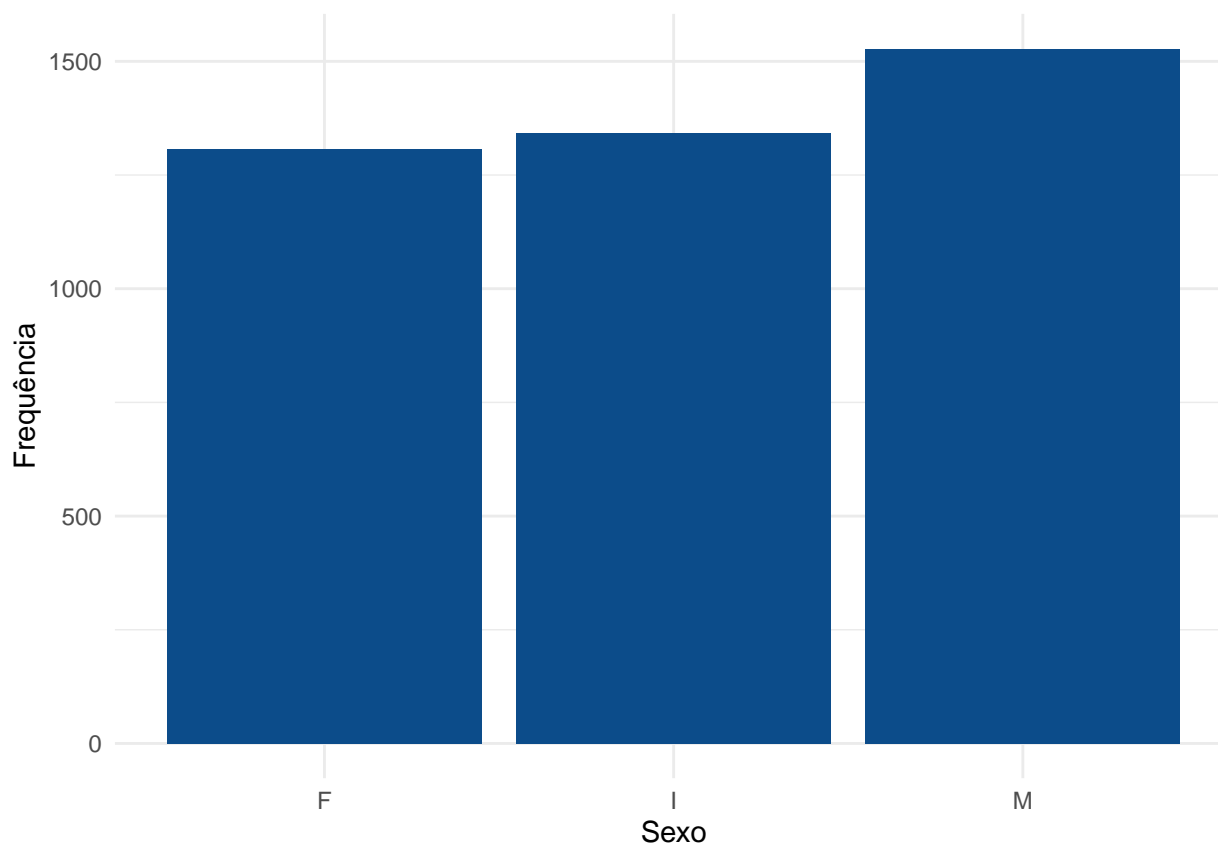
- ii. (0,5 ponto) *Elabore um gráfico de barras.*

```
library(ggplot2)
```

```
ggplot(dados) +
```



```
aes(x = sexo) +  
geom_bar(fill = "#0c4c8a") +  
labs(x = "Sexo", y = "Frequência") +  
theme_minimal()
```



4. Para as variáveis *sexo* e *aneis*, pede-se:

- i. (0,8 ponto) Construa uma tabela de contingência com perfil linha e outra com perfil coluna.

```
# Perfil linha
```

```
ctable(dados$sexo, dados$aneis, prop = "r")
```

```
## Cross-Tabulation, Row Proportions
```

```
## sexo * aneis
```

```
## Data Frame: dados
```

```
##
```

```
## -----  
##      aneis      1      2      Total  
##  sexo  
##    F      424 (32.4%)  883 (67.6%) 1307 (100.0%)  
##    I     1095 (81.6%)  247 (18.4%) 1342 (100.0%)  
##    M      577 (37.8%)  951 (62.2%) 1528 (100.0%)  
##  Total     2096 (50.2%) 2081 (49.8%) 4177 (100.0%)  
## -----
```

```
# Perfil coluna
```

```
ctable(dados$sexo, dados$aneis, prop = "c")
```

Se apenas uma das tabelas
tiver sido feita, dê apenas a
metade da questão.

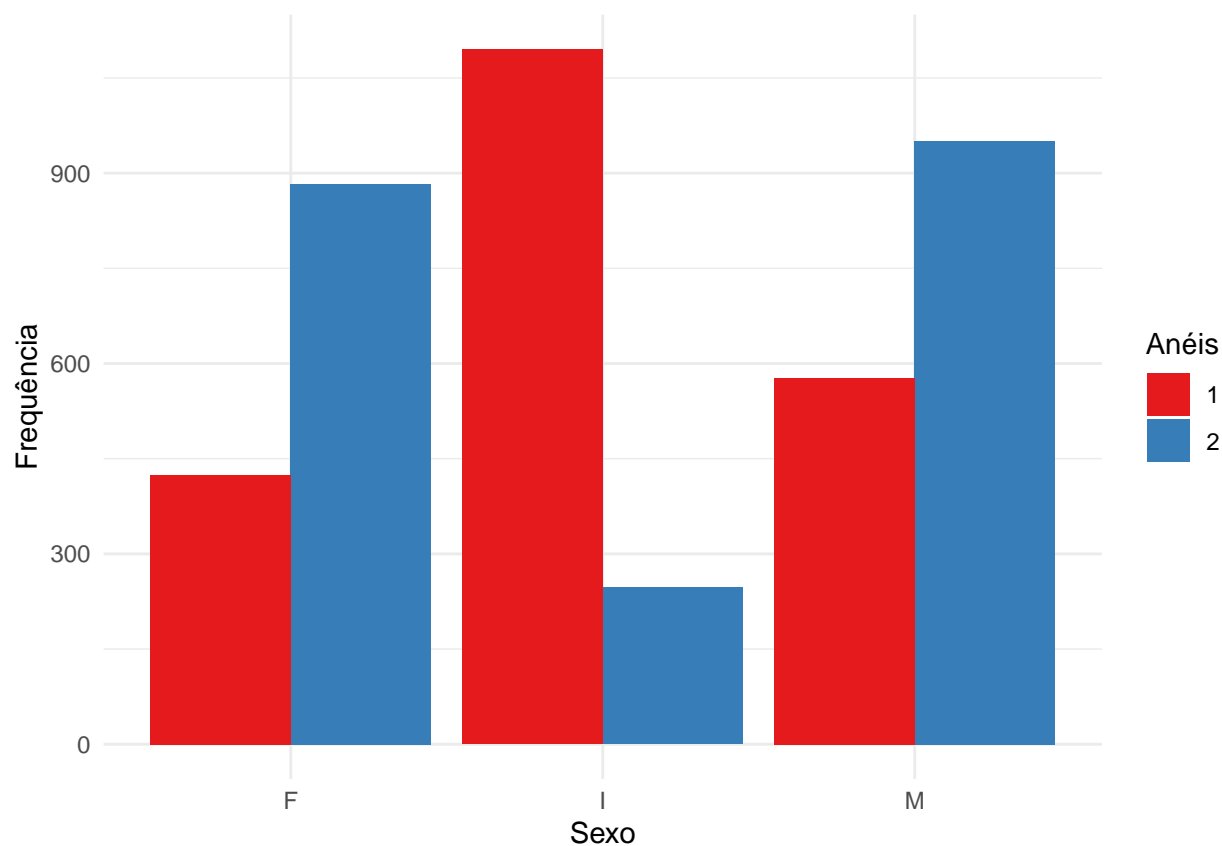


```
## Cross-Tabulation, Column Proportions
## sexo * aneis
## Data Frame: dados
##
## -----
##          aneis          1          2          Total
##  sexo
##    F          424 ( 20.2%)      883 ( 42.4%)    1307 ( 31.3%)
##    I          1095 ( 52.2%)      247 ( 11.9%)    1342 ( 32.1%)
##    M           577 ( 27.5%)      951 ( 45.7%)    1528 ( 36.6%)
##  Total          2096 (100.0%)     2081 (100.0%)    4177 (100.0%)
## -----
```

iii. (1,0 ponto) *Construa um gráfico com barras múltiplas e outro com barras empilhadas.*

```
library(ggplot2)

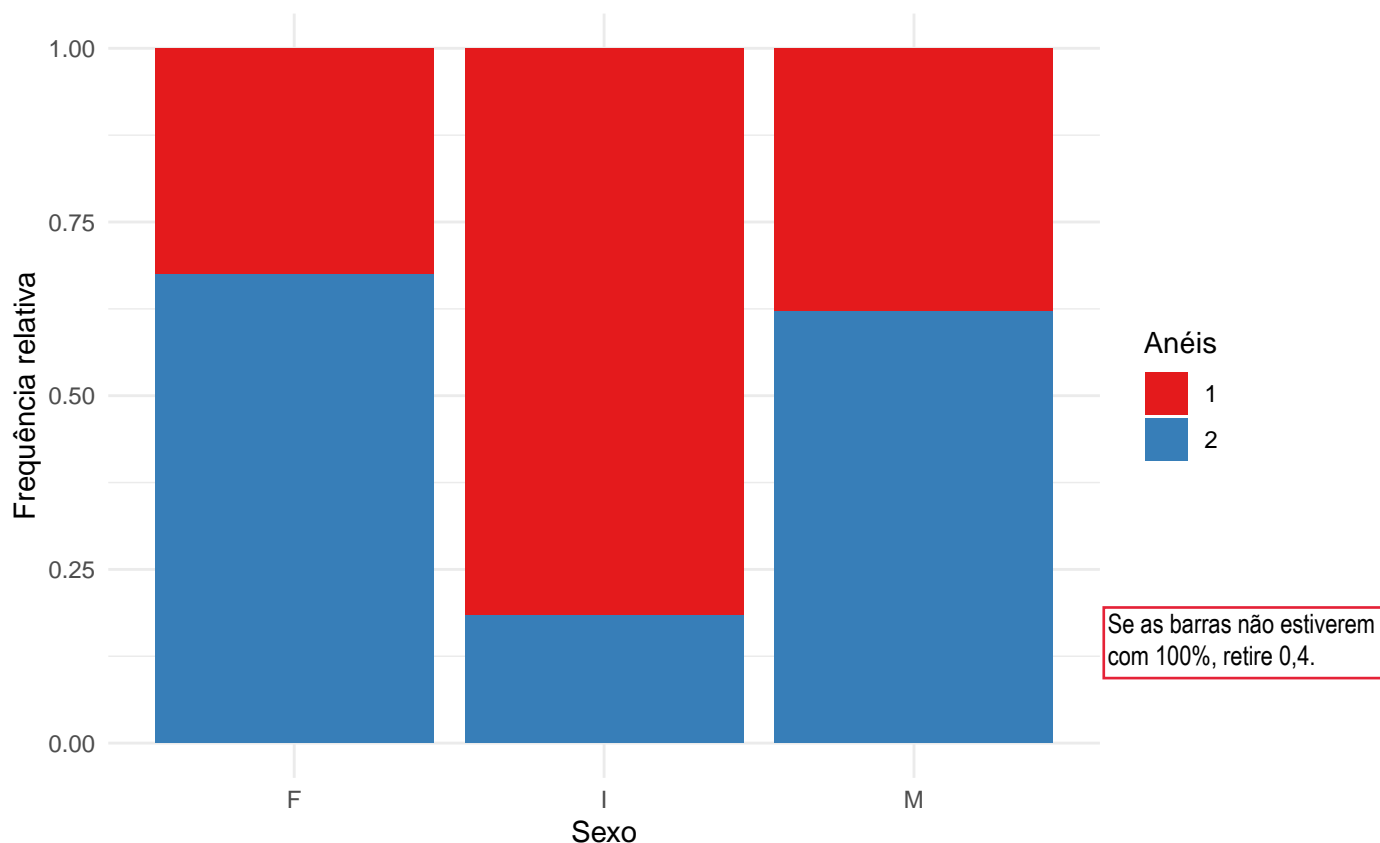
# Barras múltiplas
ggplot(dados) +
  aes(x = sexo, fill = aneis) +
  geom_bar(position = "dodge") +
  scale_fill_brewer(palette = "Set1") +
  labs(x = "Sexo", y = "Frequência", fill = "Anéis") +
  theme_minimal()
```



```
# Barras empilhadas
ggplot(dados) +
```




```
aes(x = sexo, fill = aneis) +  
geom_bar(position = "fill") +  
scale_fill_brewer(palette = "Set1") +  
labs(x = "Sexo", y = "Frequência relativa", fill = "Anéis") +  
theme_minimal()
```



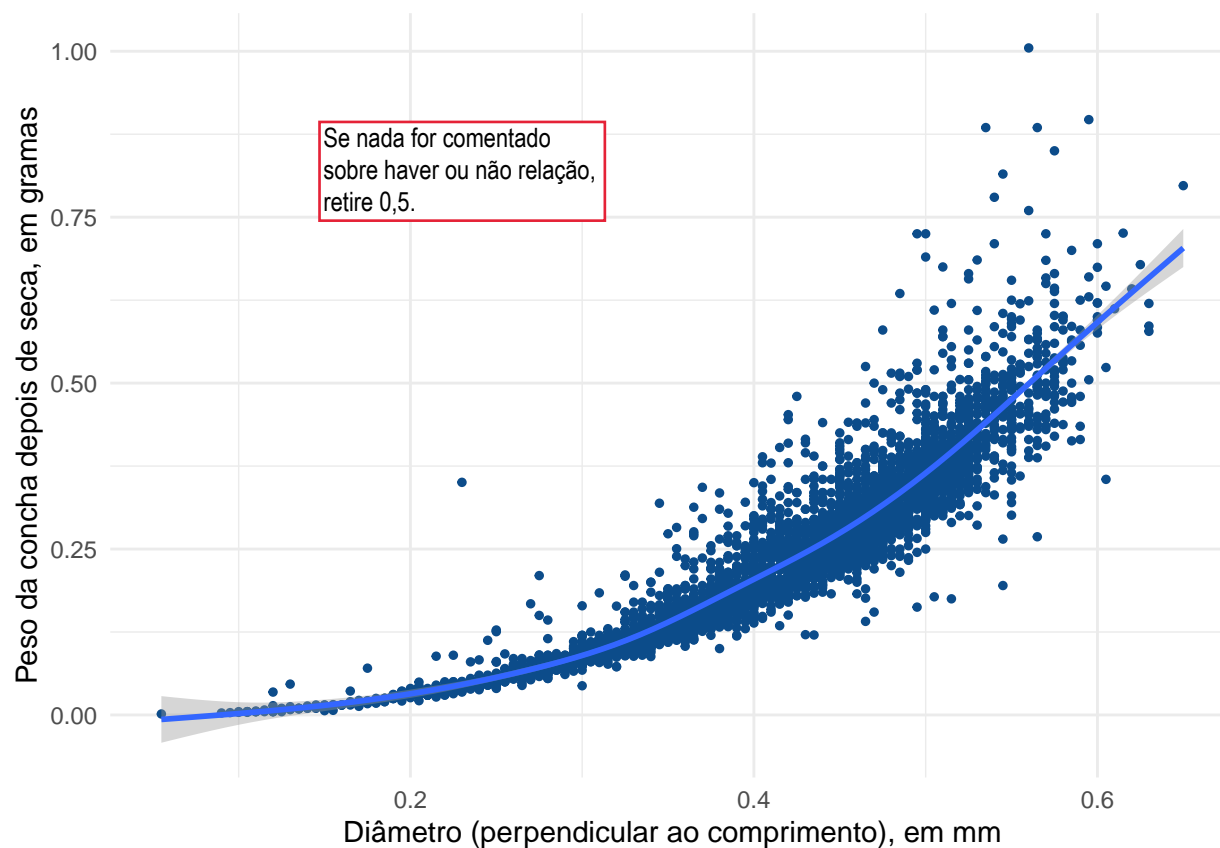
5. Para as variáveis *diâmetro* e *peso_concha*:

- i. (1,0 ponto) Construa um gráfico de dispersão e avalie se há indícios de relação entre as variáveis.

```
library(ggplot2)
```

```
ggplot(dados) +  
  aes(x = diâmetro, y = peso_concha) +  
  geom_point(size = 1L, colour = "#0c4c8a") +  
  geom_smooth(span = 0.75) +  
  labs(x = "Diâmetro (perpendicular ao comprimento), em mm", y = "Peso da concha depois de seca, em gram") +  
  theme_minimal()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

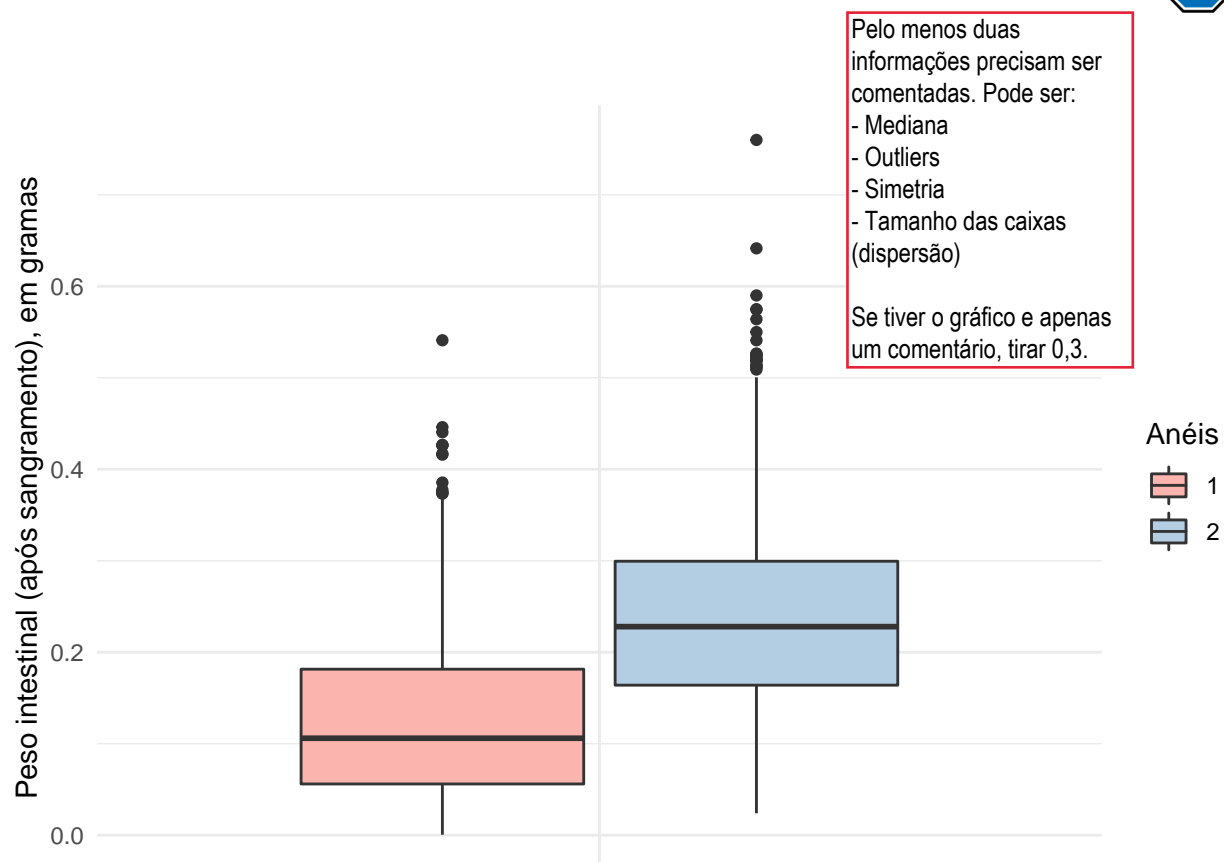


É possível ver claramente uma relação entre as variáveis. Conforme aumenta o diâmetro, também aumenta o peso da concha.

6. (1,0 ponto) Construa boxplots para a variável *peso_intestinal*, segundo a variável *aneis*. Descreva o que é observado no gráfico.

```
library(ggplot2)

ggplot(dados) +
  aes(x = "", y = peso_intestinal, fill = aneis) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Pastel1") +
  labs(x = " ", y = "Peso intestinal (após sangramento), em gramas", fill = "Anéis") +
  theme_minimal()
```



É possível observar que o grupo 2 (que possui mais de 10 anéis), apresenta uma mediana maior comparada ao grupo 1 (até 10 anéis). Além disso, ambos os grupos apresentam assimetria à direita e outliers superiores.

DESAFIO! (Bônus de 1,0 ponto) Construa um gráfico não ensinado em sala de aula utilizando alguma(s) variável(is) do banco de dados. Explique o que o gráfico mostra.