

CASO PRÁTICO 2 – Tópicos de Resolução

Estatística Descritiva, Introdução ao R

Exercício 1

Considere os dados do exercício 4 do Caso Prático 1, relativos às notas dos alunos de uma turma

9, 15, 15, 17, 14, 16, 19, 10, 14, 15, 18, 16, 13, 17

No  RStudio :

- a) crie um vector com os dados das notas;

`Notas <- c(9, 15, 15, 17, 14, 16, 19, 10, 14, 15, 18, 16, 13, 17)`

- b) obtenha a tabela de frequências;

`> table(Notas) → Tabela de frequências (absolutas)`

9 10 13 14 15 16 17 18 19
1 1 1 2 3 2 2 1 1

`prop.table(table(Notas)) → Tabela de frequências (relativas)`

Notas
9 10 13 14 15 16
0.07142857 0.07142857 0.07142857 0.14285714 0.21428571 0.14285714 0.14285714
17
0.07142857 0.07142857 0.07142857 0.07142857 0.07142857 0.07142857 0.07142857
18 19
0.07142857 0.07142857

`round(prop.table(table(Notas)), 3) → Tabela de frequências (relativas) arredonda
mento a 3 casas decimais`

Notas
9 10 13 14 15 16 17 18 19
0.071 0.071 0.071 0.143 0.214 0.143 0.143 0.071 0.071

```

prop.table(table(Notas))*100
Notas
   9      10      13      14      15      16
 7.142857 7.142857 7.142857 14.285714 21.428571 14.285714
   17      18      19
14.285714 7.142857 7.142857

```

```

round(prop.table(table(Notas))*100, 3)
Notas
   9      10      13      14      15      16      17      18
 7.143    7.143    7.143  14.286  21.429  14.286  14.286  7.143
   19
 7.143

```

- c) obtenha a média, mediana, desvio padrão, mínimo, máximo, amplitude total, quartis, quantil de 65%;

```
mean(Notas)
[1] 14.85714
```

```
median(Notas)
[1] 15
```

```
sd(Notas)
[1] 2.797173
```

```
max(Notas)
[1] 19
```

```
min(Notas)
[1] 9
```

```
diff(range(Notas)) (amplitude total)
[1] 10
```

```
quantile(Notas, 0.25)
25%
14
```

```
quantile(Notas, 0.75)
75%
16.75
```

```
quantile(Notas)
 0%   25%   50%   75% 100%
 9.00 14.00 15.00 16.75 19.00
```

```
summary(Notas)
Min. 1st Qu. Median     Mean 3rd Qu.     Max.
 9.00    14.00   15.00    14.86   16.75    19.00
```

```
quantile(Notas,.65)
65%
16
```

P65%=16

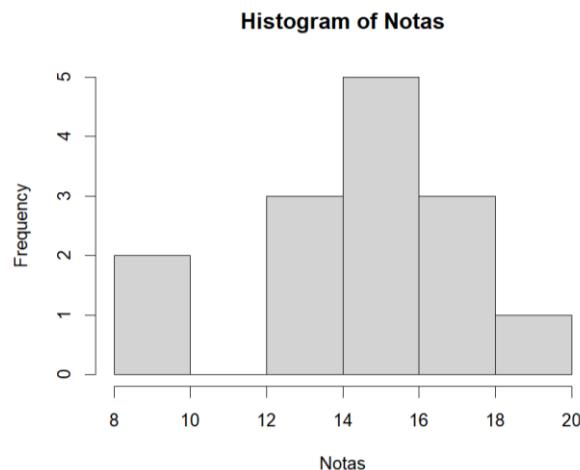
65% dos alunos tiveram nota entre 9 e 16 valores e os restantes 35% tiveram nota entre 16 e 19 valores.

`me=15`

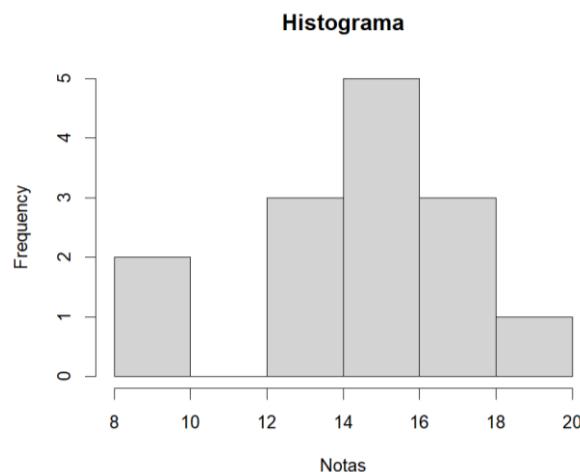
50% dos alunos tiveram nota entre 9 e 15 valores e os restantes 50% tiveram nota entre 15 e 19 valores.

d) Obtenha o histograma e a box-plot

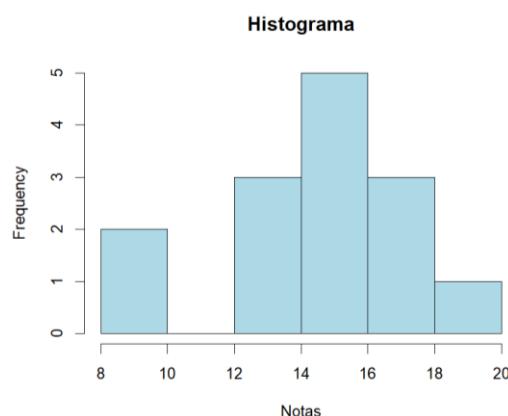
`hist(Notas)`



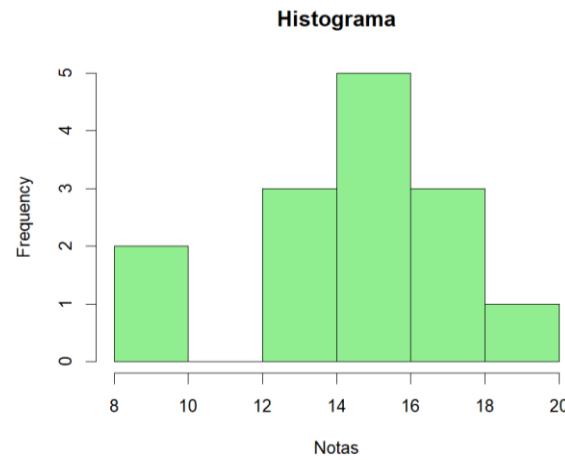
`hist(Notas, main = "Histograma")`



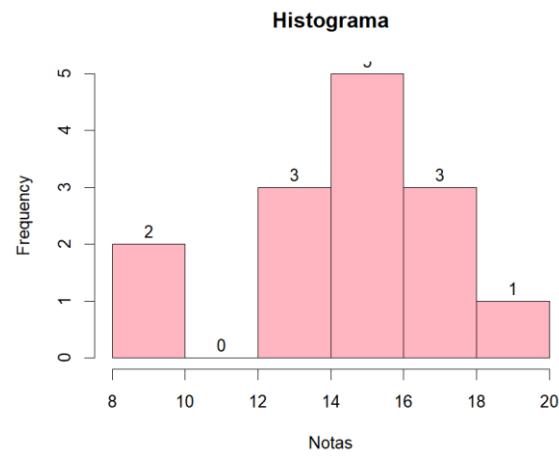
`hist(Notas, main = "Histograma", col = "lightblue")`



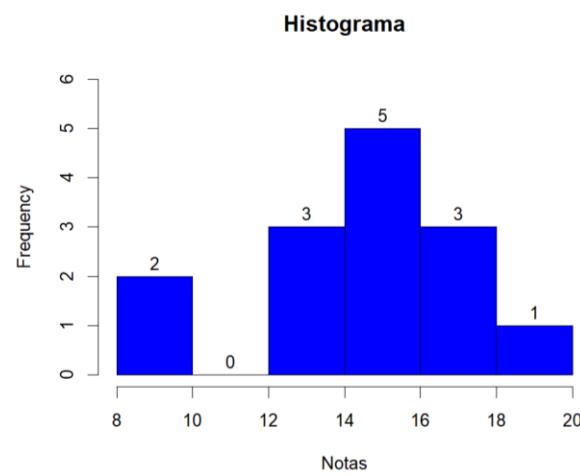
```
hist(Notas, main = "Histograma", col = "lightgreen")
```



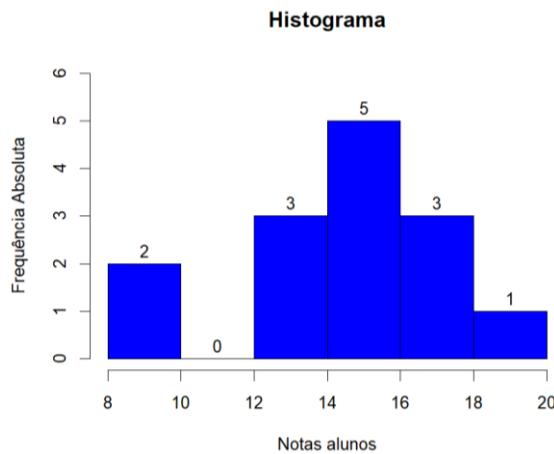
```
hist(Notas, main = "Histograma", col = "lightpink", labels = TRUE)
```



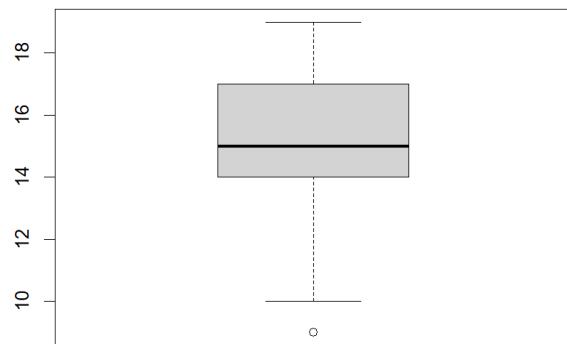
```
hist(Notas, main = "Histograma",
     col = "blue",
     labels = TRUE,
     ylim = c(0,6))
```



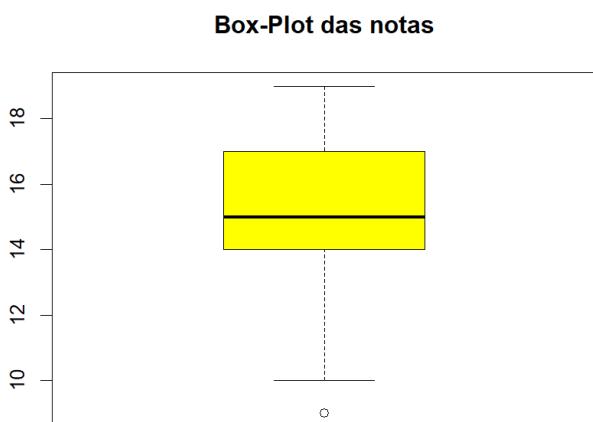
```
hist(Notas, main = "Histograma",
+     col = "blue", labels = TRUE, xlab = "Notas alunos",
+     ylab = "Frequência Absoluta", ylim = c(0,6))
```



```
boxplot(Notas)
```



```
boxplot(Notas, main = "Box-Plot das notas", col = "yellow")
```



- e) Interprete os outputs obtidos.
- As notas dos alunos situaram-se entre 9 e 19 valores.
 - A nota média foi 14,9 valores com uma dispersão de cerca de 2,8 valores.
 - 25% das notas situam-se entre 9 e 14 valores e as restantes 75% entre 14 e 19 valores;

- 75% das notas situam-se entre 9 e $16,75 \approx 17$ valores e as restantes 25% entre $16,75 \approx 17$ e 19 valores.
- 25% das notas entre 9 e 14 valores, 25% das notas entre 14 e 15 valores, 25% das notas entre 15 e $16,75 \approx 17$ valores e as restantes 25% entre $16,75 \approx 17$ e 19 valores.
- Existe um outlier (9 valores)
- Assimetria à esquerda → maior concentração das notas nos maiores valores 😊

Exercício 2

Considere as temperaturas máximas diárias registadas numa determinada região de Portugal (graus)

24.6, 25.2, 26.6, 25.9, 27.4, 28.5, 30.0, 30.4, 30.4, 28.9, 27.1, 27.8, 28.9, 29.1, 28.8, 27.2, 26.8, 27.9, 30.1, 27.1, 24.0, 23.1, 24.5, 25.2, 27.0, 27.2, 26.2, 25.6, 23.5, 24.0, 25.7

Utilizando funções do  RStudio responda às seguintes questões:

- a) Quantas temperaturas foram registadas?

```
length(temperaturas)
[1] 31
```

Foram registadas 31 temperaturas.

- b) Qual a temperatura máxima registada? E a mínima?

```
max(temperaturas)
[1] 30.4
```

```
> min(temperaturas)
[1] 23.1
```

As temperaturas variaram entre 23,1 e 30,4 graus.

- c) Qual a temperatura média?

```
mean(temperaturas)
[1] 26.92581
```

A temperatura média é quase 27 graus.

- d) A dispersão das temperaturas relativamente à temperatura média é 2,08 graus, aproximadamente.

```
sd(temperaturas)
[1] 2.081981
```

- e) Ordene a amostra.

```
sort(temperaturas)
[1] 23.1 23.5 24.0 24.0 24.5 24.6 25.2 25.2 25.6 25.7 25.9 26.2 2
6.6 26.8 27.0
[16] 27.1 27.1 27.2 27.2 27.4 27.8 27.9 28.5 28.8 28.9 28.9 29.1
30.0 30.1 30.4
```

[31] 30.4

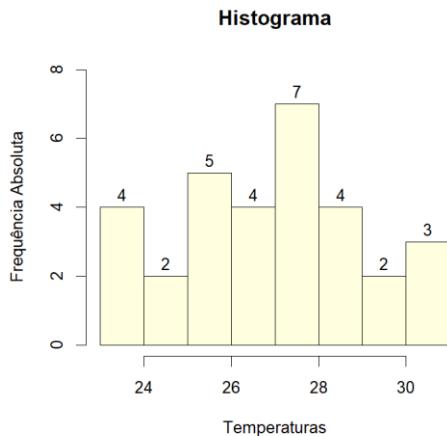
- f) Obtenha a mediana e interprete o resultado obtido.

```
median(temperaturas)  
[1] 27.1
```

50% das temperaturas situam-se entre 23,1 e 27,1 graus e as restantes temperaturas (50%) situam-se entre 27,1 e 30,4 graus.

- g) Obtenha o histograma das temperaturas e analise-o.

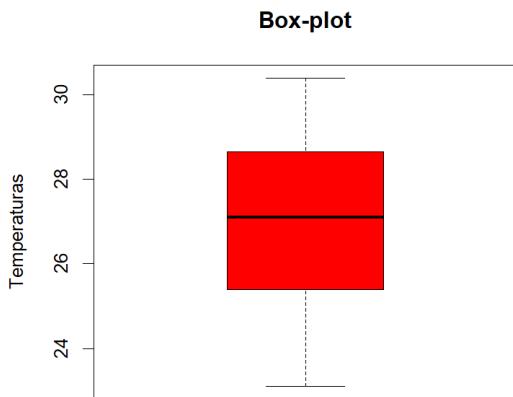
```
hist(temperaturas, main = "Histograma", labels = TRUE,  
+      ylim = c(0,8), col = "lightyellow", xlab = "Temperaturas",  
+      ylab = "Frequência Absoluta")
```



Maior concentração das temperaturas entre 25 e 29 graus.

- h) Obtenha a box-plot e analise-a.

```
boxplot(temperaturas, main = "Box-plot", col = "red", ylab = "Temperaturas")
```



```
summary(temperaturas)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
temperaturas	23.10	25.40	27.10	26.93	28.65	30.40

As temperaturas situam-se entre 23,1 e 30,4 graus.

25% das temperaturas situam-se entre 23,1 e 25,4 graus e as restantes 75% entre 25,4 e 0,04 graus.

50% das temperaturas situam-se entre 23,1 e 27,1 graus e as restantes 50% entre 27,1 e 30,04 graus.

75% das temperaturas situam-se entre 23,1 e 28,65 graus e as restantes 25% entre 28,65 e 30,04 graus.

Atendendo à Box-plot e às medidas de localização central, a distribuição das temperaturas é quase simétrica (maior concentração nos valores centrais).

- i) Relativamente à simetria/assimetria, a distribuição das temperaturas é quase simétrica ou ligeiramente assimétrica à direita.

Média=26.92581 < mediana=27.1 → ligeira assimetria à esquerda

A mesma indicação analisando a Box-plot.

- j) Obtenha a tabela de frequências das temperaturas.

```
table(temperaturas)
temperaturas
23.1 23.5 24 24.5 24.6 25.2 25.6 25.7 25.9 26.2 26.6 26.8 27
27.1 27.2
  1   1    2    1    1    2    1    1    1    1    1    1    1
  2   2
27.4 27.8 27.9 28.5 28.8 28.9 29.1
  1   1    1    1    1    2    1    30 30.1 30.4
  1   1
```

- k) Agrupe os dados das temperaturas em 4 classes.

```
table(cut(temperaturas, seq(23, 31, 1 = 5)))
```

(23,25]	(25,27]	(27,29]	(29,31]
6	9	11	5

Registaram-se

- 6 temperaturas entre 23 e 25 graus;
- 9 entre 25 e 27 graus;
- 11 entre 27 e 29 graus;
- 5 entre 29 e 31 graus.

- l) Agrupe os dados das temperaturas em 5 classes.

```
table(cut(temperaturas, seq(20, 32, 1 = 6)))
```

(20,22.4]	(22.4,24.8]	(24.8,27.2]	(27.2,29.6]	(29.6,32]
0	6	13	8	4

```
table(cut(temperaturas, seq(21, 32, 1 = 6)))
```

(21,23.2]	(23.2,25.4]	(25.4,27.6]	(27.6,29.8]	(29.8,32]
1	7	12	7	4

