ESTATÍSTICA I – LISTA DE EXERCÍCIOS 1
GABARITO

1. Para o conjunto de dados abaixo, determine:

   Produtividade da cultura da soja (kg por hectare)

3600	3545	3658	3498	3657	3425	3785	3254	3266	3641
3687	3698	3621	3654	3554	3569	3598	3578	3567	3574

   (a) Classificação da variável.
   (b) A média.
   (c) A variância.
   (d) O coeficiente de variação.

   Resposta

   (a) Quantitativa contínua.
   (b) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{71429}{20} = 3571.45$
   (c) $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \frac{327402.95}{20-1} = 17231.734$
   (d) $CV = \frac{S}{\bar{X}} \times 100 = \frac{\sqrt{17231.734}}{3571.45} = 3.676\%$

2. Os dados em rol relacionados a seguir referem-se à produção diária de leite de vacas da raça Holandesa obtida em duas ordenhas, em kg.

4.0	4.5	5.0	5.0	5.0	5.5	6.0	6.0	6.5	6.5	6.5
6.5	7.0	7.0	7.0	7.0	7.0	7.0	7.5	8.5	9.0	9.0
9.0	9.5	10.0	10.0	10.5	10.5	11.0	12.0	12.5	13.0	13.0

   Calcule o coeficiente de variação desse conjunto de dados.

   Resposta

   O coeficiente de variação é calculado da seguinte forma:

```
> media <- mean(leite)
> desvio <- sd(leite)
> media
```

```
[1] 8

> desvio

[1] 2.537223

> desvio / media * 100 # coeficiente de variação

[1] 31.71529
```

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{264}{33} = 8$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}} = \sqrt{\frac{206}{33-1}} = 2.537$$

$$CV = \frac{S}{\bar{X}} \times 100 = \frac{2.537}{8} \times 100 = 31.715\%$$

3. A tabela seguinte apresenta a produção de café, em milhões de toneladas, na região DELTA.

	Ano	Ton.
1	1992	12
2	1993	15
3	1994	18
4	1995	22
5	1996	17
6	1997	14
7	1998	18
8	1999	23
9	2000	29
10	2001	12

(a) Calcule o valor da produção média.

(b) Calcule o valor da mediana da produção.

(c) Calcule o valor do desvio padrão da produção.

Resposta

```
> mean(cafe) # média

[1] 18

> median(cafe) # mediana

[1] 17.5

> sd(cafe) # desvio-padrão

[1] 5.374838
```

(a) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{180}{10} = 18$

(b) $Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(5)} + X_{(6)}}{2} = \frac{17+18}{2} = 17.5$

(c) $S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}} = \sqrt{\frac{260}{10-1}} = \sqrt{28.889} = 5.375$

4. Foi realizado na região Oeste do Paraná, no Município de Marechal Cândido Rondon, em 1992, um levantamento da produtividade leiteira diária de 30 produtores rurais, atendidos pelo plano "Panela Cheia" (Roesler, 1997). Os resultados da produção diária dos 30 produtores estão apresentados a seguir:

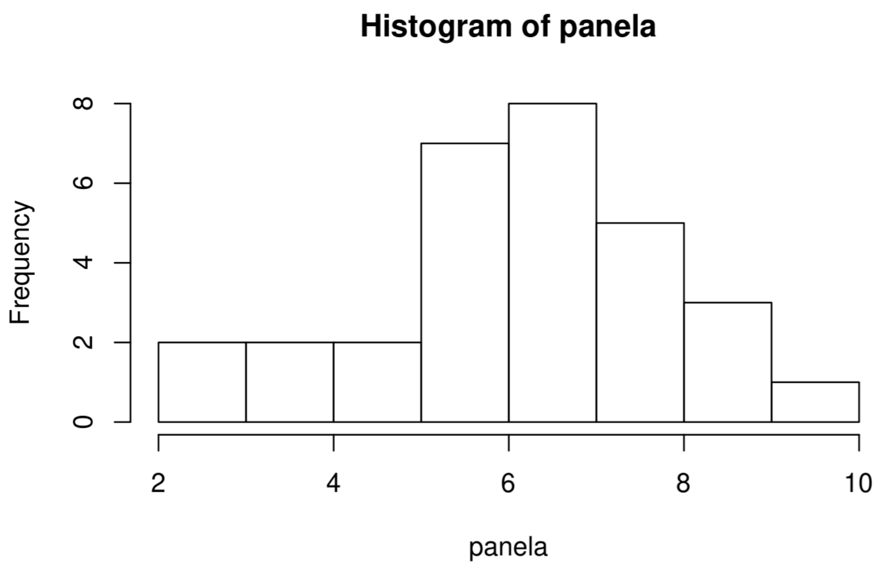
3.83	4.08	4.59	5.80	7.81	6.31	8.37	7.50	9.30	5.98
6.78	5.23	7.50	6.62	2.80	6.51	5.27	5.44	6.08	6.66
8.49	3.86	2.91	5.82	7.47	6.52	6.61	7.80	5.62	8.23

Faça uma representação gráfica para os dados.

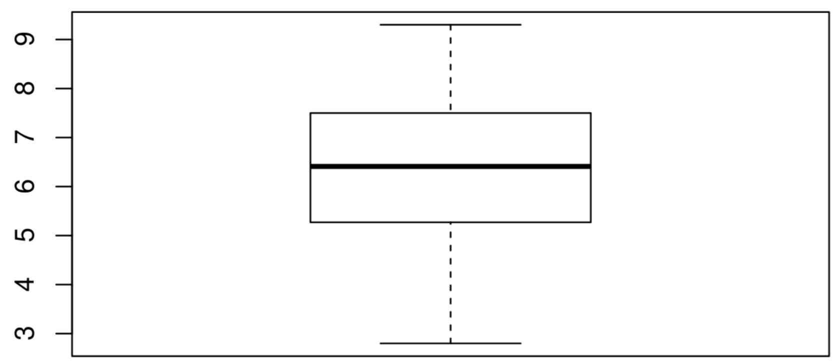
Resposta

Serão apresentadas duas alternativas de resposta:

```
> hist(panela)
```



> boxplot(panela)



5. Um talhão de 3 hectares de cana-de-açúcar foi subdividido em parcelas de 1000 m^2 cada uma. As produções dessas parcelas, em toneladas, foram as que se seguem:

9.3	7.8	8.3	10.1	10.2	9.5	8.7	9.0	8.7	9.7
9.1	8.8	3.6	9.4	3.6	8.9	9.2	9.4	11.4	3.1
9.6	3.1	2.0	9.8	8.7	9.0	8.6	9.2	10.1	9.3

(a) Calcule o valor da produção média.

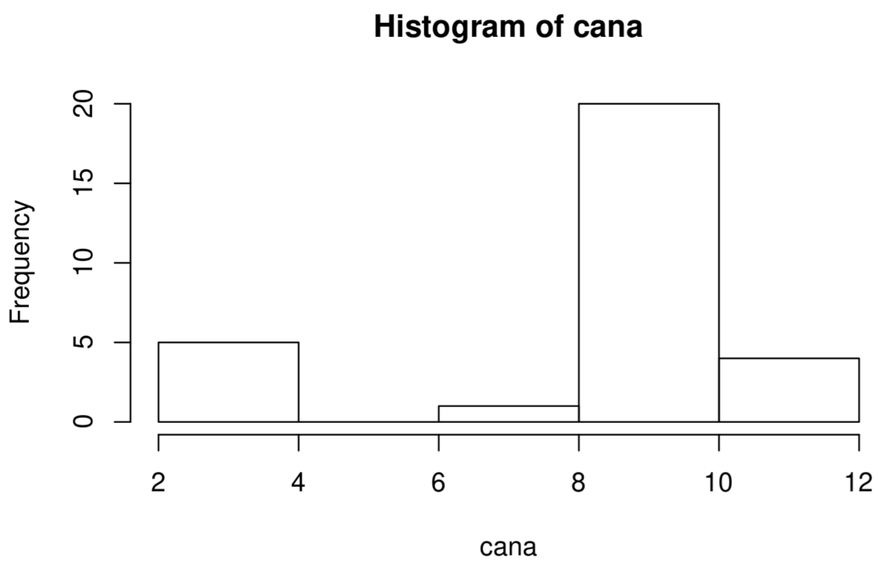
(b) Calcule o valor da mediana da produção.

(c) Faça uma representação gráfica para o conjunto de dados.

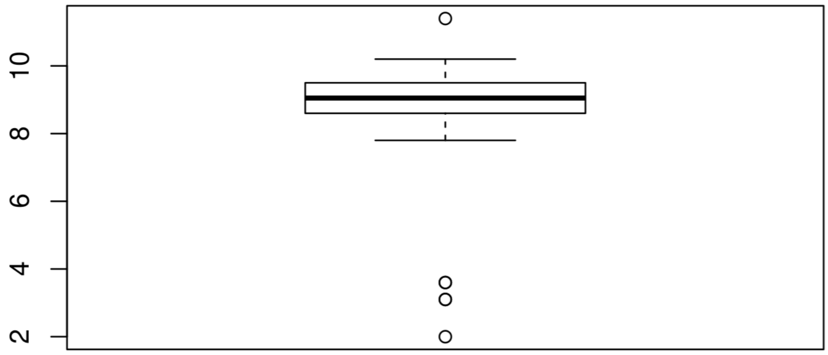
(d) Compare os valores da média e da mediana e explique a diferença encontrada.

Resposta

```
> hist(cana)
```




```
> boxplot(cana)
```



> *mean(cana)*

[1] 8.24

> *median(cana)*

[1] 9.05

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{247.2}{30} = 8.24$$

$$Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(15)} + X_{(16)}}{2} = \frac{9 + 9.1}{2} = 9.05$$

Diferença entre a média e a mediana é um indicativo de assimetria no conjunto de dados, que também pode ser verificada nos gráficos apresentados. No caso, foram verificados alguns talhões com resultados muito baixos, que faz com que o valor da média também seja puxado para baixo. Esse efeito não é verificado no cálculo da mediana que verifica apenas o valor central do conjunto de dados.

6. (Magalhães pg 28 – adaptado) O índice de germinação dos principais fatores para definir a qualidade das sementes. Ele é determinado em experimento cientifico conduzido pelo fabricante e regulamentado pelos órgãos fiscalizadores. Um fabricante afirma que o índice de germinação de suas sementes de milho é de 85%. Para verificar

tal afirmação, uma cooperativa de agricultores sorteou 20 amostras com 100 sementes em cada uma e anotou a porcentagem de germinação em cada amostra.

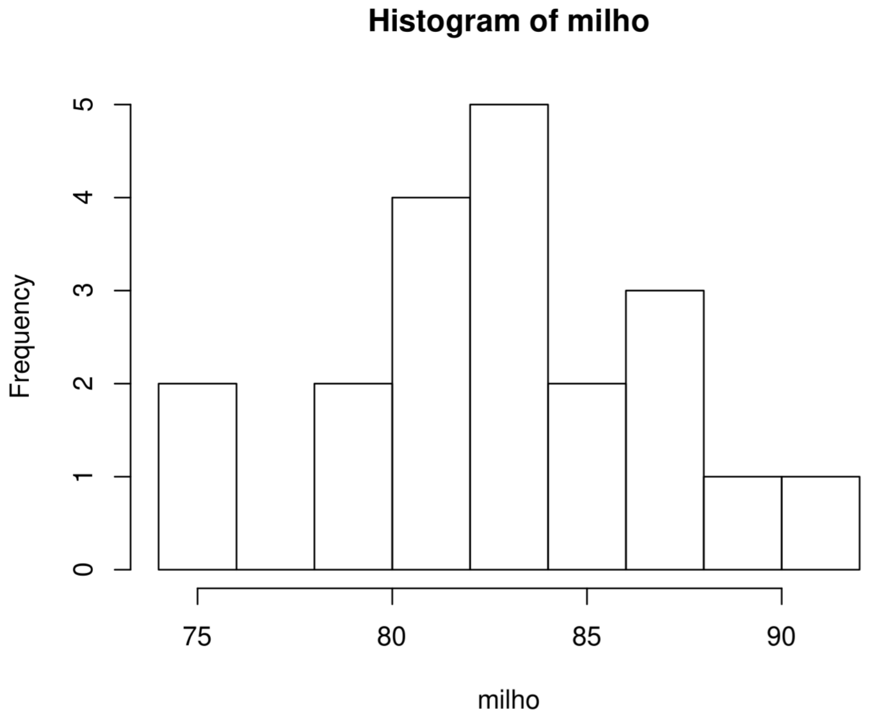
82.1	74.9	80.4	85.3	90.5	82.4	85.1	82.7	75.4	80.7
80.8	82.7	89.1	87.2	82.1	81.6	86.8	86.1	79.2	79.1

(a) Faça uma representação gráfica da tabela acima.

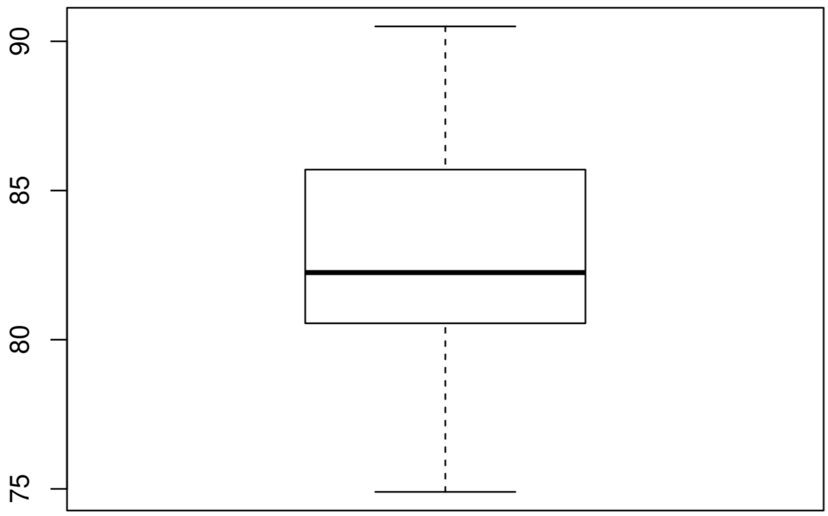
(b) Comente a afirmação do fabricante.

Resposta

```
> hist(milho)
```




```
> boxplot(milho)
```




```
> mean(milho)

[1] 82.71
```

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1654.2}{20} = 82.71$$

Se verifica que a média da amostra foi menor que o índice informado pelo fabricante. Porém, como a diferença foi pequena, é difícil afirmar que a média da população de sementes é realmente menor que o informado, uma vez que foram analisadas apenas 20 amostras. Testes estatísticos são indicados para se avaliar essa diferença.

7. (Magalhães pg 28 – adaptado) Uma nova ração foi fornecida a suínos recém desmamados e deseja se avaliar sua eficiência. A ração tradicional dava um ganho de peso ao redor de 3,5 kg em um mês. A seguir, apresentamos os dados referentes o ganho, em quilos, para essa nova ração, aplicada durante um mês em 20 animais nas condições acima.

2.94	3.38	2.49	3.52	2.97	2.09	2.91	1.74	4.27	5.17
2.27	1.79	3.16	2.47	5.99	2.55	3.29	2.61	1.99	2.76

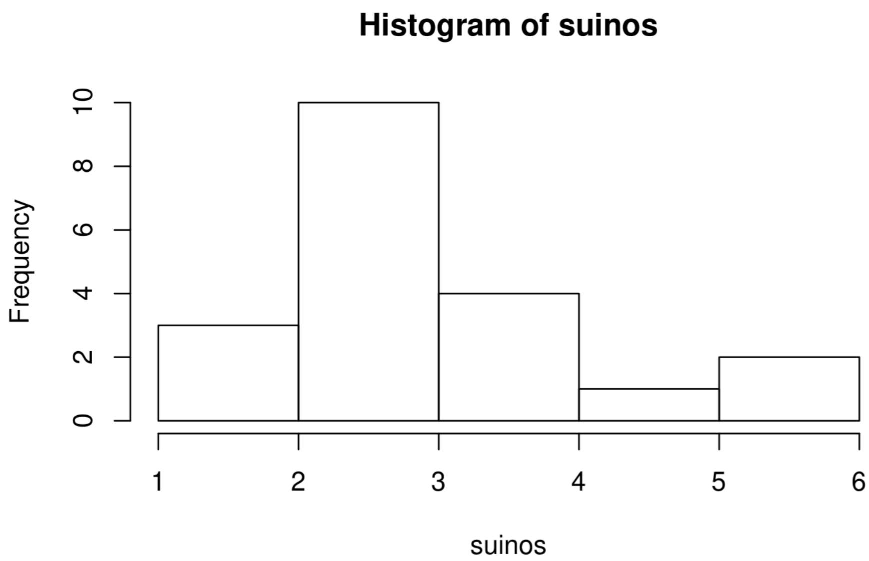
(a) Construa o histograma.

(b) Determine o 1º, 2º e 3º quartis.

(c) Você acha que a nova ração é mais eficiente que a tradicional? Justifique.

Resposta

```
> hist(suinos)
```




```
> mean(suinos)
```

```
[1] 3.018
```

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{60.36}{20} = 3.018$$

```
> quantile(suinos, c(0.25, 0.50, 0.75), type=2)
```

```
  25%   50%   75%
2.370 2.835 3.335
```

$$Q_1 = \frac{X_{\left(\frac{n}{4}\right)} + X_{\left(\frac{n}{4}+1\right)}}{2} = \frac{X_{(5)} + X_{(6)}}{2} = \frac{2.27 + 2.47}{2} = 2.37$$

$$Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(10)} + X_{(11)}}{2} = \frac{2.76 + 2.91}{2} = 2.835$$

$$Q_3 = \frac{X_{\left(\frac{3n}{4}\right)} + X_{\left(\frac{3n}{4}+1\right)}}{2} = \frac{X_{(15)} + X_{(16)}}{2} = \frac{3.29 + 3.38}{2} = 3.335$$

Apenas verificando os resultados já se nota que o ganho médio de peso na amostra foi muito menor que o da ração tradicional, que indica que a nova ração na verdade é menos eficiente.

8. A tabela abaixo se refere ao número de dias consecutivos sem chuva em algumas cidades de uma região do sertão da Paraíba.

	Intervalo	fi	Fi	fr	Fr
1	[8,10)		3	10.0	
2	[10,12)	9	12	30.0	40.0
3	[12,14)	7			
4	[14,16)			13.3	76.7
5	[16,18)	4	27		90.0
6	[18,20]	3	30	10.0	100.0

(a) Qual a classificação da variável?

(b) Complete a tabela para encontrar a média de dias sem chuva.

(c) Encontre também a variância.

Resposta

	Intervalo	fi	Fi	fr	Fr
1	[8,10)	3	3	10.0	10.0
2	[10,12)	9	12	30.0	40.0
3	[12,14)	7	19	23.3	63.3
4	[14,16)	4	23	13.3	76.7
5	[16,18)	4	27	13.3	90.0
6	[18,20]	3	30	10.0	100.0

(a) Quantitativa discreta

(b) $\bar{X} = \frac{\sum_{i=1}^{n}(X_i \times f_i)}{\sum_{i=1}^{n} f_i} = \frac{27+99+91+60+68+57}{30} = \frac{402}{30} = 13.4$

(c) $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \times f_i}{n-1} = \frac{58.08+51.84+1.12+10.24+51.84+94.08}{30-1} = \frac{267.2}{29} = 9.21$

9. Em uma granja foi observado o peso de 80 frangos:

961	967	971	974	979	982	992	996	997	998
998	999	1000	1000	1002	1002	1002	1003	1004	1004
1005	1005	1007	1007	1008	1009	1009	1009	1010	1013
1014	1016	1016	1017	1017	1020	1021	1021	1022	1024
1025	1026	1027	1028	1028	1028	1029	1030	1030	1030
1030	1031	1031	1032	1032	1032	1033	1034	1034	1034
1038	1038	1039	1040	1041	1041	1041	1042	1049	1052
1054	1056	1065	1069	1071	1072	1079	1082	1083	1084

Queremos dividir os frangos em 4 categorias, com relação ao peso, de modo que:

i. Os 20% mais leves sejam da categoria A.
ii. Os 30% seguintes sejam da categoria B.
iii. Os 30% seguintes sejam da categoria C.
iv. Os 20% seguintes (ou seja, os 20% mais pesados) sejam da categoria D.

Quais os limites de peso entre as categorias A, B, C e D?

Resposta

Estamos interessados em encontrar os quantis 20% (segundo decil), 50% (mediana) e 80% (oitavo decil).

```
> quantis <- c(0, 0.2, 0.5, 0.8, 1)
> quantile(frangos, quantis, type=2)
```

```
   0%    20%    50%    80%   100%
961.0 1002.0 1024.5 1040.5 1084.0
```

$$D_2 = \frac{X_{\left(\frac{n}{5}\right)} + X_{\left(\frac{n}{5}+1\right)}}{2} = \frac{X_{(16)} + X_{(17)}}{2} = \frac{1002 + 1002}{2} = 1002$$

$$Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(40)} + X_{(41)}}{2} = \frac{1024 + 1025}{2} = 1024.5$$

$$D_8 = \frac{X_{\left(\frac{4n}{5}\right)} + X_{\left(\frac{4n}{5}+1\right)}}{2} = \frac{X_{(64)} + X_{(65)}}{2} = \frac{1040 + 1041}{2} = 1040.5$$

Categoria	Limites
A	$961 \vdash 1002$
B	$1002 \vdash 1024.5$
C	$1024.5 \vdash 1040.5$
D	$1040.5 \vdash 1084$

10. Foi verificado em duas localidades o crescimento de árvores de uma determinada espécie em um intervalo de tempo definido:

Localidade 1	9.4	9.2	10.2	12.5	12.2	10.5	12.1	10.2
	11.4	10.1	10.1	8.7	8.5	6.5	9.6	
Localidade 2	10.8	11.7	11.5	11.7	11	11.9	11.3	11.7
	10.5	10.7	11.6	12.4	11.8	12.4	12.5	

Fonte: software R, pacote *agricolae*, base de dados *growth*.

Há algum indício de que a localidade em que a árvore está plantada causou influência em seu crescimento?

Resposta

```
> require(agricolae) # pacote que contém a base de dados
> data(growth) # armazena a base de dados
> attach(growth) # separa as colunas da tabela
```

Média por localidade.

```
> tapply(height, place, mean)

      L1       L2
10.08000 11.56667
```

Variância por localidade.

```
> tapply(height, place, var)

      L1       L2
2.504571 0.382381
```

Desvio-padrão por localidade.

```
> tapply(height, place, sd)
```

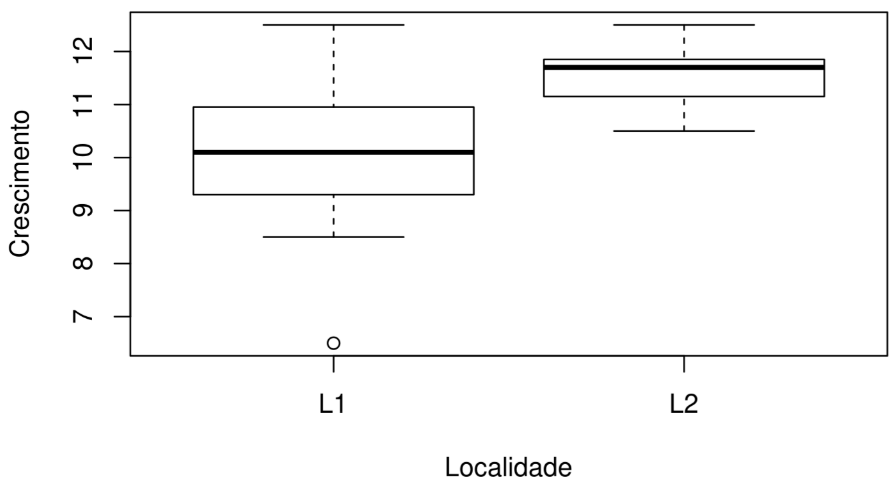
```
      L1        L2
1.5825838 0.6183696
```

Coeficiente de variação por localidade.

```
> tapply(height, place, sd) / tapply(height, place, mean) * 100
```

```
      L1        L2
15.700236  5.346135
```

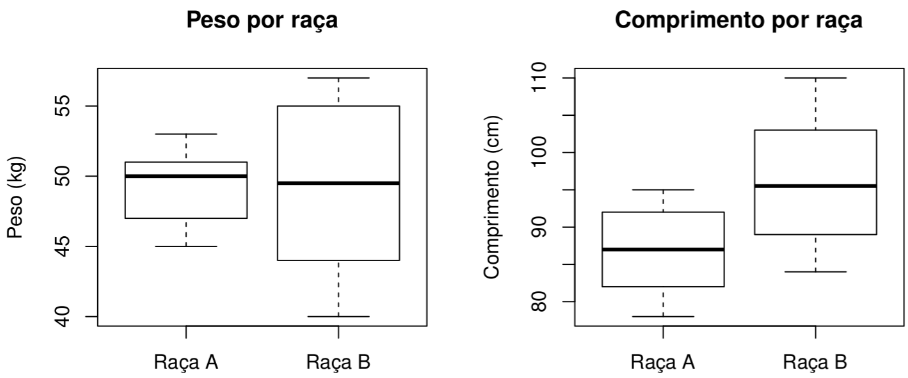
```
> boxplot(height ~ place, xlab="Localidade", ylab="Crescimento")
```




```
> detach(growth) # remove os dados da memória
```

De acordo com os resultados, o crescimento das árvores foi um pouco maior na localidade L2, e como a dispersão dos dados foi pequena, este é um forte indicativo de que essa diferença verificada entre as duas localidades é realmente significativa.

11. Para os resultados abaixo referentes a pesos e comprimentos de bezerros de um determinado confinamento:



Peso

	n	Média	Variância	CV (%)
Raça A	10	49.4	6.933333	5.330207
Raça B	10	49.4	36.488889	12.227942

Comprimento

	n	Média	Variância	CV (%)
Raça A	10		39.6	7.266574
Raça B	10	95.8		8.647028

Os valores omitidos da tabela são respectivamente:

(a) 65.4 e 11.079

(b) 72.7 e 11.079

(c) 86.6 e 68.622

(d) 86.6 e 8.284

(e) 65.4 e 68.622

Resposta

Para o primeiro valor ausente

$$CV = \frac{S}{\bar{X}} \times 100$$

$$7.266574 = \frac{\sqrt{39.6}}{\bar{X}} \times 100$$

$$\bar{X} = \frac{\sqrt{39.6}}{7.266574} \times 100$$

$$\bar{X} = 86.6$$

Para o segundo valor ausente

$$CV = \frac{S}{\bar{X}} \times 100$$

$$8.647028 = \frac{S}{95.8} \times 100$$

$$S = \frac{8.647028 \times 95.8}{100}$$

$$S = 8.283853 \qquad \rightarrow \qquad S^2 = 68.622222$$

(a) FALSO
(b) FALSO
(c) VERDADEIRO
(d) FALSO
(e) FALSO

12. Utilizando os dados da Tabela 1

Tabela 1: Informações sobre estado civil, grau de instrução, nº de filhos, salário (expresso como fração do salário mínimo), idade (medida em anos) e procedência de 36 funcionários da seção de orçamentos, da Companhia Mista.

	Estado civil	Grau de instrução	Nº filhos	Salário	Idade	Região de procedência
1	Solteiro	Ens. Fund.		4,00	26	Interior
2	Casado	Ens. Fund.	1	4,56	32	Capital
3	Casado	Ens. Fund.	2	5,25	36	Capital
4	Solteiro	Ens. Médio		5,73	20	Outro
5	Solteiro	Ens. Fund.		6,26	40	Outro
6	Casado	Ens. Fund.	0	6,66	28	Interior
7	Solteiro	Ens. Fund.		6,86	41	Interior
8	Solteiro	Ens. Fund.		7,39	43	Capital
9	Casado	Ens. Médio	1	7,59	34	Capital
10	Solteiro	Ens. Médio		7,44	23	Outro
11	Casado	Ens. Médio	2	8,12	33	Interior
12	Solteiro	Ens. Fund.		8,46	27	Capital
13	Solteiro	Ens. Médio		8,74	37	Outro
14	Casado	Ens. Fund.	3	8,95	44	Outro
15	Casado	Ens. Médio	0	9,13	30	Interior
16	Solteiro	Ens. Médio		9,35	38	Outro
17	Casado	Ens. Médio	1	9,77	31	Capital
18	Casado	Ens. Fund.	2	9,80	39	Outro
19	Solteiro	Superior		10,53	25	Interior
20	Solteiro	Ens. Médio		10,76	37	Interior
21	Casado	Ens. Médio	1	11,06	30	Outro
22	Solteiro	Ens. Médio		11,59	34	Capital
23	Solteiro	Ens. Fund.		12,00	41	Outro
24	Casado	Superior	0	12,79	26	Outro
25	Casado	Ens. Médio	2	13,23	32	Interior
26	Casado	Ens. Médio	2	13,60	35	Outro
27	Solteiro	Ens. Fund.		13,85	46	Outro
28	Casado	Ens. Médio	0	14,69	29	Interior
29	Casado	Ens. Médio	5	14,71	40	Interior
30	Casado	Ens. Médio	2	15,99	35	Capital
31	Solteiro	Superior		16,22	31	Outro
32	Casado	Ens. Médio	1	16,61	36	Interior
33	Casado	Superior	3	17,26	43	Capital
34	Solteiro	Superior		18,75	33	Capital
35	Casado	Ens. Médio	2	19,40	48	Capital
36	Casado	Superior	3	23,30	42	Interior

Tabela: Estatística Básica, Wilton O. Bussab e Pedro A. Morettim, pg 4.
Fonte: dados hipotéticos.

(a) Construa a distribuição de frequência absoluta, frequência relativa, porcentagem e porcentagem acumulada, para as variáveis estado civil, grau de instrução, número de filhos e região de procedência.

(b) Construa um gráfico de pizza para a variável grau de instrução e um para região de procedência.

(c) Construa um gráfico de barras para a variável estado civil e um para numero de filhos.

(d) Construa um histograma para a variável salário e um para idades.

(e) A média da variável número de filhos, a mediana da variável idade o primeiro e terceiro quartil da variável salário.

Resposta

```
> read.table("tabela.txt", header=TRUE)
> attach(tabela)
```

Tabela para estado civil.

```
> tabcivil <- data.frame(fi = as.numeric(table(civil)))
> tabcivil$Fi <- cumsum(table(civil))
> tabcivil$fr <- table(civil) / length(civil) * 100
> tabcivil$Fr <- cumsum(table(civil)) / length(civil) * 100
> rownames(tabcivil) <- names(table(civil))
> tabcivil
```

```
         fi Fi       fr        Fr
Casado   20 20 55.55556  55.55556
Solteiro 16 36 44.44444 100.00000
```

Tabela para grau de instrução.

```
> tabgrau <- data.frame(fi = as.numeric(table(grau)))
> tabgrau$Fi <- cumsum(table(grau))
> tabgrau$fr <- table(grau) / length(grau) * 100
> tabgrau$Fr <- cumsum(table(grau)) / length(grau) * 100
> rownames(tabgrau) <- names(table(grau))
> tabgrau
```

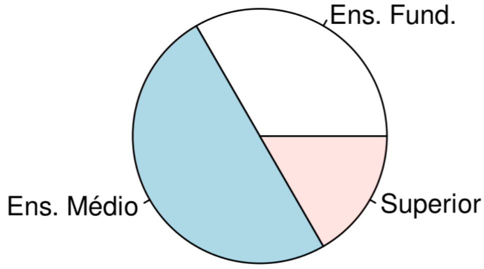
```
           fi Fi       fr        Fr
Ens. Fund. 12 12 33.33333  33.33333
Ens. Médio 18 30 50.00000  83.33333
Superior    6 36 16.66667 100.00000
```

Tabela para número de filhos.

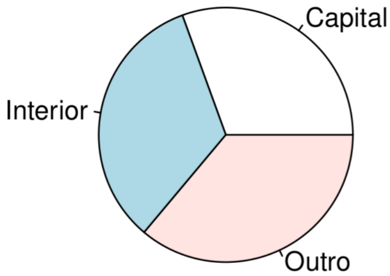
```
> tabfilhos <- data.frame(fi = as.numeric(table(filhos)))
> tabfilhos$Fi <- cumsum(table(filhos))
> tabfilhos$fr <- table(filhos) / length(filhos[complete.cases(filhos)]) * 100
> tabfilhos$Fr<-cumsum(table(filhos))/length(filhos[complete.cases(filhos)])*100
> rownames(tabfilhos) <- names(table(filhos))
> tabfilhos

  fi Fi fr  Fr
0  4  4 20  20
1  5  9 25  45
2  7 16 35  80
3  3 19 15  95
5  1 20  5 100
```

Tabela para região de procedência.

```
> tabregiao <- data.frame(fi = as.numeric(table(regiao)))
> tabregiao$Fi <- cumsum(table(regiao))
> tabregiao$fr <- table(regiao) / length(regiao) * 100
> tabregiao$Fr <- cumsum(table(regiao)) / length(regiao) * 100
> rownames(tabregiao) <- names(table(regiao))
> tabregiao

         fi Fi       fr        Fr
Capital  11 11 30.55556  30.55556
Interior 12 23 33.33333  63.88889
Outro    13 36 36.11111 100.00000

> pie(table(grau), labels=names(table(grau)))
```

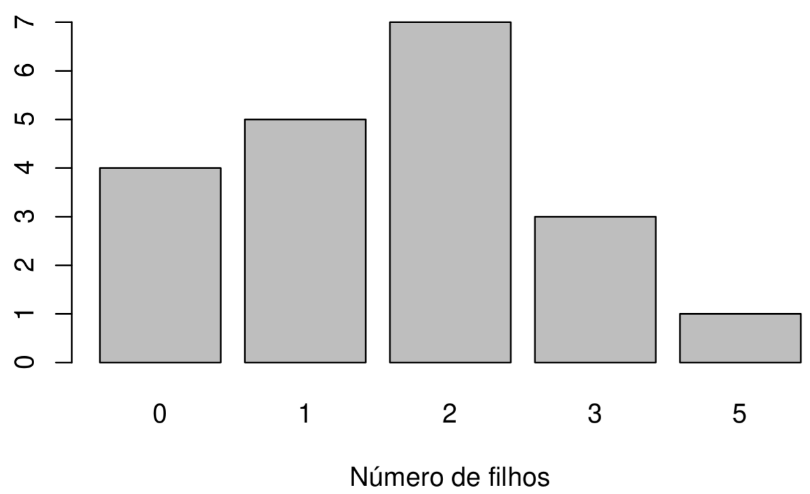



```
> pie(table(regiao), labels=names(table(regiao)))
```

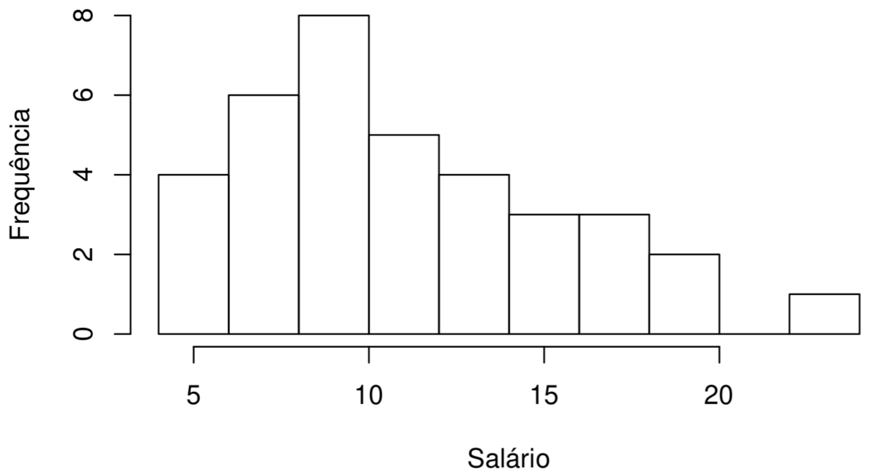



```
> barplot(table(civil), xlab="Estado civil")
```

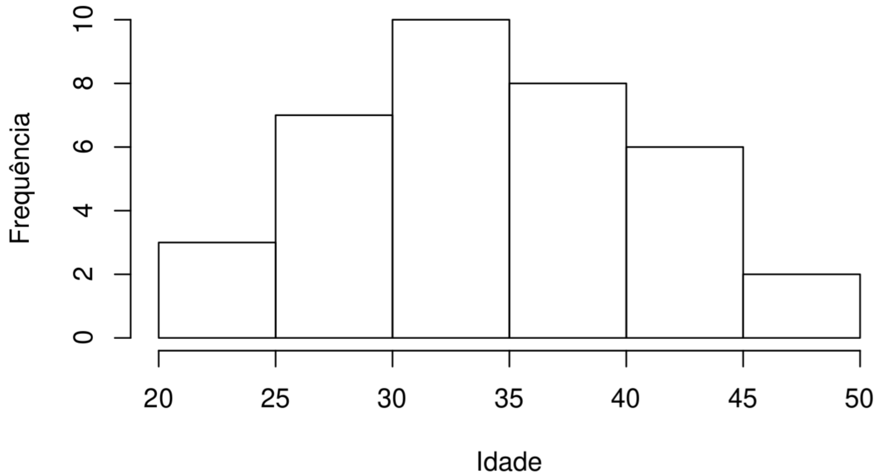

```
> barplot(table(filhos), xlab="Número de filhos")
```




```
> hist(salario, main=NULL, xlab="Salário", ylab="Frequência")
```




```
> hist(idade, main=NULL, xlab="Idade", ylab="Frequência")
```




```
> mean(filhos, na.rm=TRUE)
[1] 1.65
> median(idade)
```

```
[1] 34.5

> quantile(salario, c(0.25, 0.75), type=2)

   25%    75%
 7.515 14.270

> detach(tabela)
```