

Noções de Probabilidade e Estatística

Resolução dos Exercícios Ímpares

CAPÍTULO 1

Felipe E. Barletta Mendes

8 de outubro de 2007

Neste Capítulo serão mostrados os comandos do R utilizados para resolver os Exercícios e suas respectivas saídas.

Exercícios da Sessão 1.1

1. Classifique em Verdadeiro ou Falso as seguintes afirmações
 - (a) Estatística é um conjunto de técnicas destinadas a organizar um conjunto de valores numéricos.
Falso
 - (b) Sempre que estivermos trabalhando com números, deveremos utilizar a Inferência Estatística.
Falso
 - (c) A Estatística Descritiva fornece uma maneira adequada de tratar um conjunto de valores, numéricos ou não, com a finalidade de conhecermos o fenômeno de interesse.
Verdadeiro
 - (d) Qualquer amostra representa, de forma adequada, uma população.
Falso
 - (e) As técnicas Estatísticas não são adequadas para casos que envolvam experimentos destrutivos como, por exemplo, queima de equipamentos, etc.
Falso

Exercícios da Sessão 1.2

- 1 Classifique cada uma das variáveis abaixo em qualitativa (nominal/ordinal) ou quantitativa (discreta/contínua).
 - (a) Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (sim ou não são possíveis respostas para esta variável).
Qualitativa Nominal
 - (b) Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos, além de não sei)
Qualitativa Nominal

- (c) Perda de peso de maratonistas na Corrida de São Silvestre, em quilos.
Quantitativa contínua
- (d) Intensidade da perda de peso maratonistas na Corrida de São Silvestre (leve, moderada, forte).
Qualitativa Ordinal
- (e) Grau de satisfação de da população brasileira em relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
Qualitativa Ordinal

3 Os dados abaixo referem-se o salário de 20 pessoas em uma indústria.

Primeiro entramos com os dados no R.

```
> dados <- c(10.1, 7.3, 8.5, 5, 4.2, 3.1, 2.2, 9, 9.4, 6.1, 3.3,
+ 10.7, 1.5, 8.2, 10, 4.7, 3.5, 6.5, 8.9, 6.1)
```

- (a) Construa uma tabela de frequência agrupando os dados em intervalos de amplitude 2 a partir de 1.

Os comandos abaixo podem auxiliar na construção da tabela

```
> table(dados)

dados
 1.5  2.2  3.1  3.3  3.5  4.2  4.7   5  6.1  6.5  7.3  8.2  8.5  8.9   9  9.4
  1   1   1   1   1   1   1   1   2   1   1   1   1   1   1   1
10 10.1 10.7
  1   1   1

> round(prop.table(table(dados)), digits = 3)

dados
 1.5  2.2  3.1  3.3  3.5  4.2  4.7   5  6.1  6.5  7.3  8.2  8.5  8.9   9  9.4
0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.10 0.05 0.05 0.05 0.05 0.05 0.05 0.05
10 10.1 10.7
0.05 0.05 0.05

> fir <- matrix(c("1 a 3", "3 a 5", "5 a 7", "7 a 9", "9 a 11",
+ "Total", 2, 5, 4, 4, 5, 20, "10%", "25%", "20%", "20%", "25%",
+ "100%", "10%", "35%", "55%", "75%", "100%", "-"), ncol = 4,
+ nrow = 6)
> dimnames(fir) <- list(c("", "", "", "", "", ""), c("Classe",
+ "Freq. Abs.", "Freq. Rel.", "Freq.Ac.))
> fir

Classe  Freq. Abs. Freq. Rel. Freq.Ac.
"1 a 3"  "2"      "10%"    "10%"
"3 a 5"  "5"      "25%"    "35%"
"5 a 7"  "4"      "20%"    "55%"
"7 a 9"  "4"      "20%"    "75%"
"9 a 11" "5"      "25%"    "100%"
"Total"  "20"     "100%"   "-"
```

A tabela de frequência pode ser construída e organizada no Latex utilizando os resultados obtidos no R usando o comando barra Sexpr .

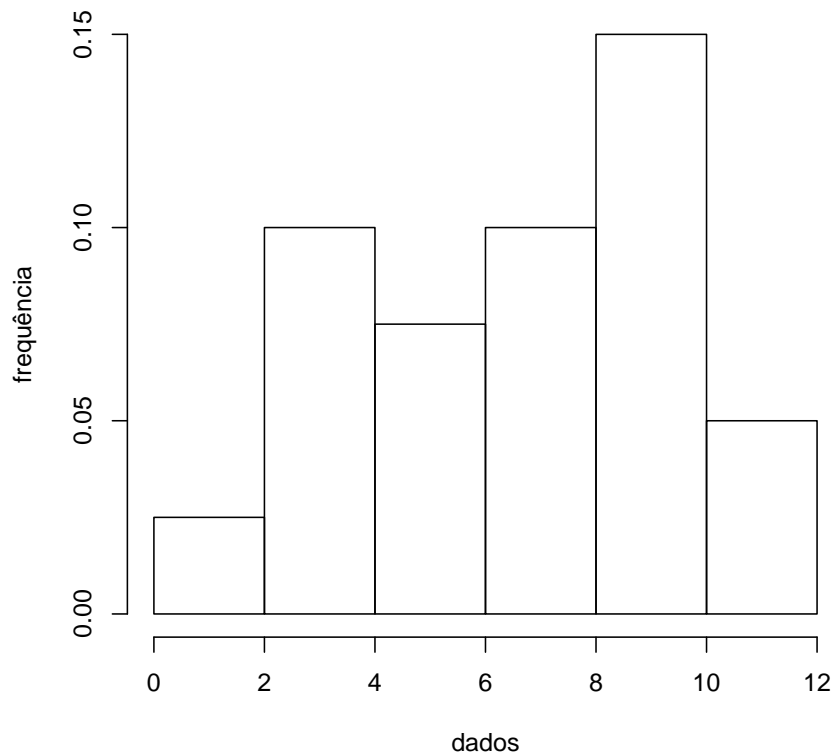
salário	Freq. Absoluta	Freq. Relativa	Freq. Acumulada
1 - 3	2	10%	10%
3 - 5	5	25%	35%
5 - 7	4	20%	55%
7 - 9	4	20%	75%
9 - 11	5	25%	100%
Total	20	100%	

(b) Construa o histograma e calcule o 1^o e 3^o quartis.

```
> hist(dados, main = "Histograma do exercício 3", ylab = "frequência",
+      prob = T)
> summary(dados)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
1.500  4.025  6.300  6.415  8.925 10.700
```

Histograma do exercício 3



5 Um estudo pretende verificar se o problema da desnutrição em adultos medida pelo peso, em quilos, em uma região agrícola(A) é maior do que uma região industrial(B). Para tanto, uma amostra foi tomada em cada região, fornecendo a tabela de frequência a seguir:

```
> menor.quarenta <- seq(20, 40, l = 8)
> quar.cinq <- seq(40, 49, l = 25)
```

Tabela 1: região A

Peso	n_i
< 40	8
40 † 50	25
50 † 60	28
60 † 70	12
≥ 70	9
Total	82

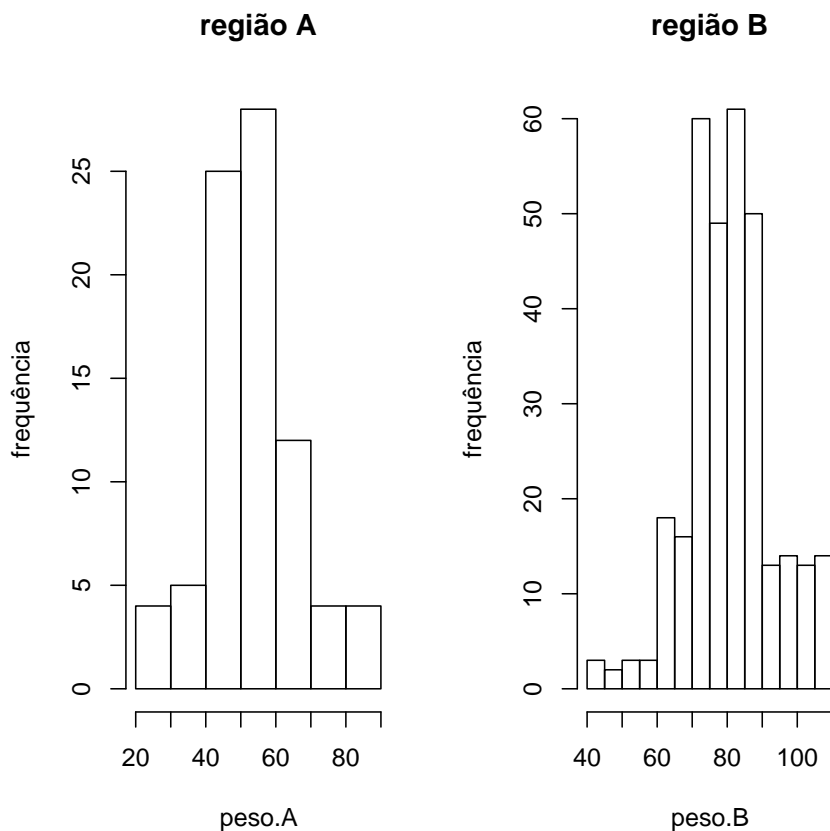
Tabela 2: região B

Peso	n_i
< 60	10
60 † 70	34
70 † 80	109
80 † 90	111
≥ 90	55
Total	319

```

> cinq.sess <- seq(50, 59, l = 28)
> sess.set <- seq(60, 69, l = 12)
> maior.set <- seq(70, 90, l = 9)
> peso.A <- c(menor.quarenta, quar.cinq, cinq.sess, sess.set, maior.set)
> menor.ssessenta <- seq(40, 59, l = 10)
> s.s <- seq(60, 69, l = 34)
> s.o <- seq(70, 79, l = 109)
> o.n <- seq(80, 89, l = 111)
> maior.n <- seq(90, 110, l = 55)
> peso.B <- c(menor.ssessenta, s.s, s.o, o.n, maior.n)
> par(mfrow = c(1, 2))
> hist(peso.A, main = "região A", ylab = "frequência")
> hist(peso.B, main = "região B", ylab = "frequência")

```



```

> summary(peso.A)

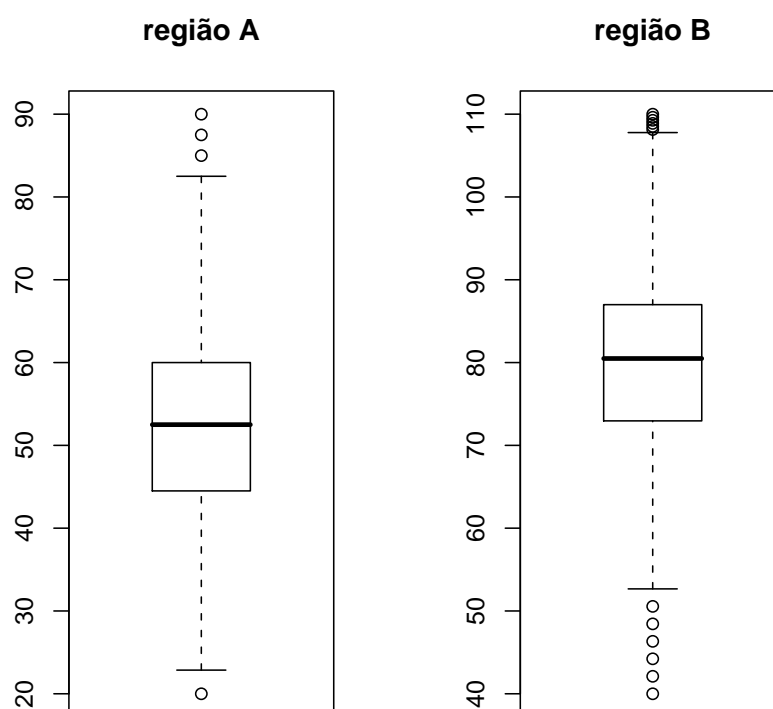
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 20.00  44.59   52.50   53.32  59.75   90.00

> summary(peso.B)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 40.00  72.96   80.49   80.53  87.00  110.00

> par(mfrow = c(1, 2))
> boxplot(peso.A, main = "região A")
> boxplot(peso.B, main = "região B")

```



Exercícios da Sessão 1.3

- 1 Com as informações da Tabela 1.1, construa um banco de dados para os 20 indivíduos iniciais e as 4 primeiras colunas.

```

> Id <- c(1:20)
> Turma <- rep("A", 20)
> Sexo <- c("F", "F", "M", "M", "F", "M", "F", "F", "F", "F", "F", "F",
+          "F", "F", "M", "F", "F", "F", "M", "F", "F")
> Idade <- c(17, 18, 18, 25, 19, 19, 20, 18, 18, 17, 18, 18, 21,

```

```
+      19, 18, 19, 17, 18, 20, 18)
> db <- data.frame(Id, Turma, Sexo, Idade)
> db
```

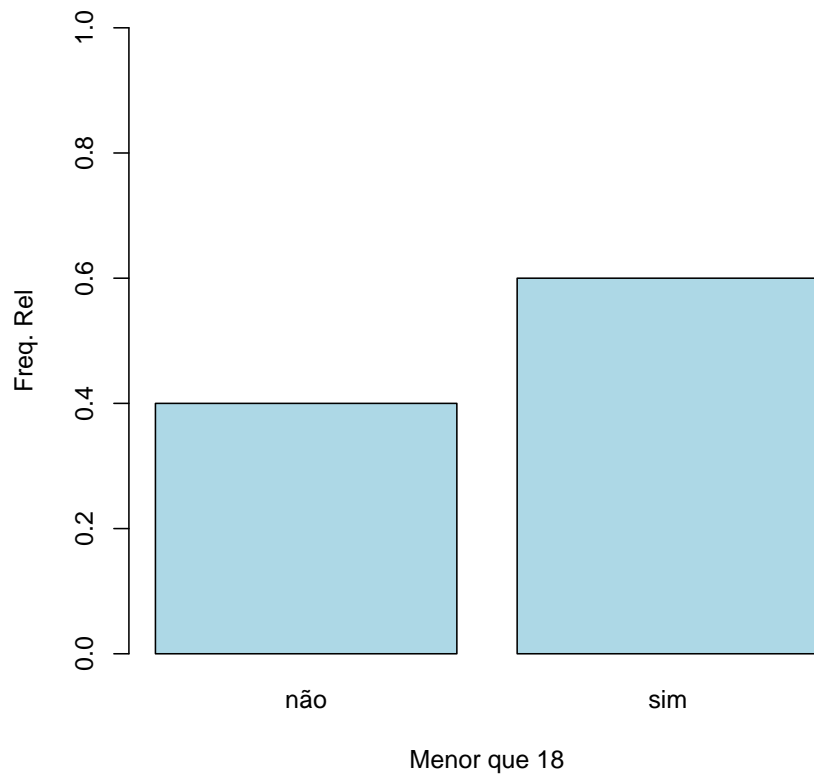
	Id	Turma	Sexo	Idade
1	1	A	F	17
2	2	A	F	18
3	3	A	M	18
4	4	A	M	25
5	5	A	F	19
6	6	A	M	19
7	7	A	F	20
8	8	A	F	18
9	9	A	F	18
10	10	A	F	17
11	11	A	F	18
12	12	A	F	18
13	13	A	F	21
14	14	A	M	19
15	15	A	F	18
16	16	A	F	19
17	17	A	F	17
18	18	A	M	18
19	19	A	F	20
20	20	A	F	18

- 3 Considerando o banco de dados do exercício 1, divida a idade em três categorias (menores de 18 anos, idade entre 18 e 21 inclusive, e maiores de 21). Construa gráficos de barra para essa variável, incluindo todos os indivíduos e um para cada sexo.

```
> Id.f <- Idade <= 18
> Id.s <- Idade > 18 & Idade <= 21
> Id.t <- Idade > 21
> menor <- factor(Id.f, labels = c("não", "sim"))
> table(menor)

menor
não sim
  8  12

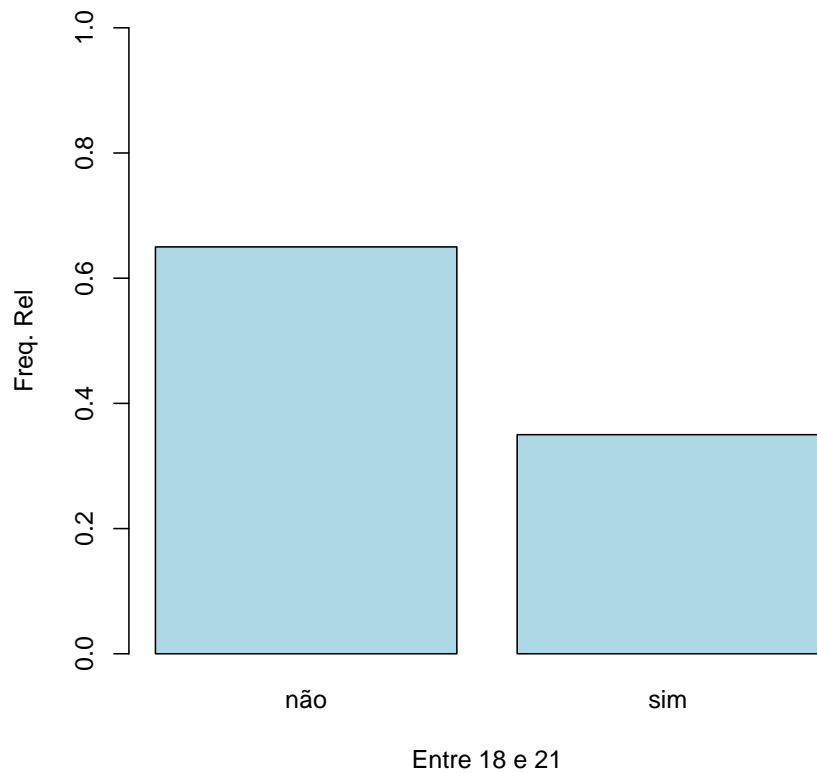
> tab.menor <- round(prop.table(table(menor)), 2)
> barplot(tab.menor, col = c("lightblue"), xlab = "Menor que 18",
+         ylim = c(0, 1), ylab = "Freq. Rel")
```



```
> entre <- factor(Id.s, labels = c("não", "sim"))
> table(entre)

entre
não sim
 13   7

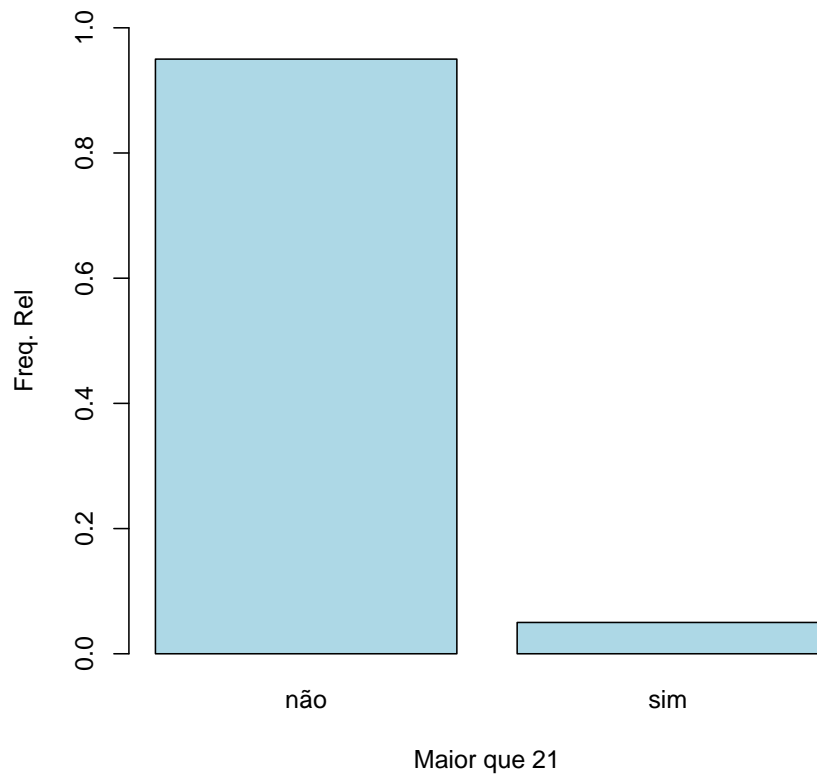
> tab.entre <- round(prop.table(table(entre)), 2)
> barplot(tab.entre, col = c("lightblue"), xlab = "Entre 18 e 21",
+         ylim = c(0, 1), ylab = "Freq. Rel")
```



```
> maior <- factor(Id.t, labels = c("não", "sim"))
> table(maior)

maior
não sim
 19  1

> tab.maior <- round(prop.table(table(maior)), 2)
> barplot(tab.maior, col = c("lightblue"), xlab = "Maior que 21",
+         ylim = c(0, 1), ylab = "Freq. Rel")
```

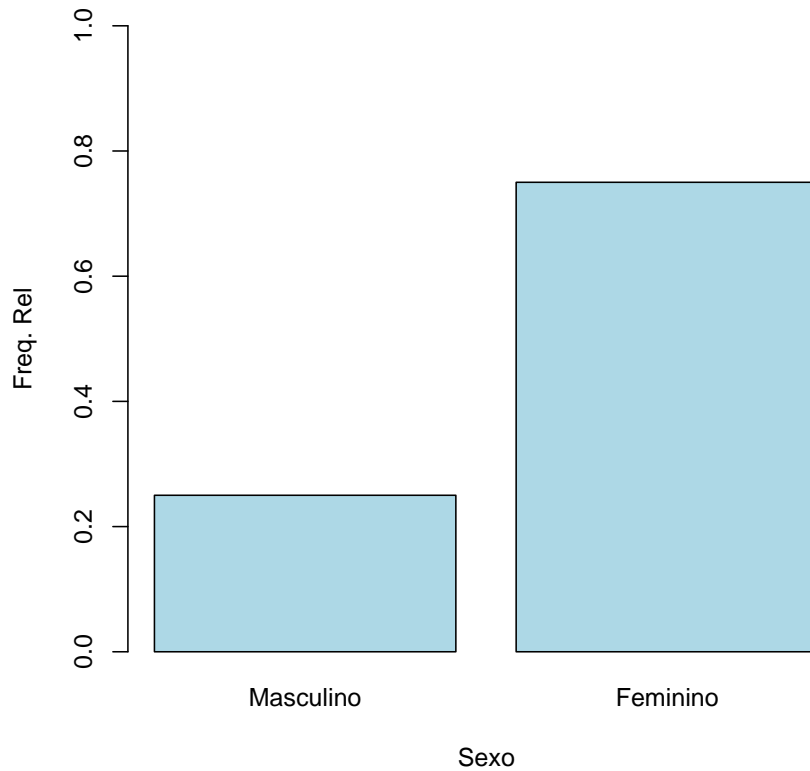
```

> sexo.f <- Sexo == "F"
> sexo.f <- factor(sexo.f, labels = c("Masculino", "Feminino"))
> table(sexo.f)

sexo.f
Masculino Feminino
         5         15

> tab.f <- round(prop.table(table(sexo.f)), 2)
> barplot(tab.f, col = c("lightblue"), xlab = "Sexo", ylim = c(0,
+ 1), ylab = "Freq. Rel")

```



Exercícios da Sessão 1.4

3 Uma pesquisa com usuários de transporte coletivo na cidade de São Paulo indagou sobre os diferentes tipos usados. Dentre ônibus, metrô e trem, o número de diferentes meios de transporte utilizados foi o seguinte:

```
> tipo <- c(2, 3, 2, 1, 2, 1, 2, 1, 2, 3, 1, 1, 1, 2, 2, 3, 1,
+          1, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 2, 3)
```

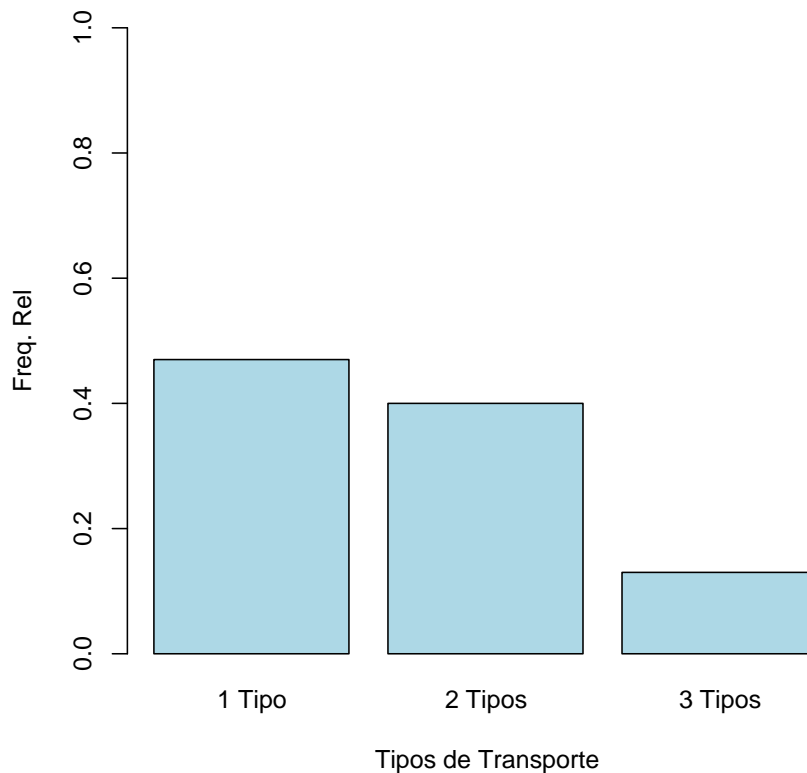
(a) Organize uma tabela de frequência.

```
> tipo.t <- factor(tipo, labels = c("1 Tipo", "2 Tipos", "3 Tipos"))
> freq.abs <- table(tipo.t)
> tab.t <- round(prop.table(table(tipo.t)), 2)
```

Nº Transp.	1	2	3	Total
frequência Abs.	14	12	4	30
frequência Rel.	0.47	0.4	0.13	1.00

(b) Faça uma representação gráfica.

```
> barplot(tab.t, col = c("lightblue"), xlab = "Tipos de Transporte",
+         ylim = c(0, 1), ylab = "Freq. Rel")
```



(c) Admitindo que essa amostra represente bem o comportamento do usuário paulistano, você acha que a porcentagem dos usuário que utilizam mais de um tipo de transporte é grande?

Sim, pois se olharmos na tabela de frequência percebemos que há mais de 50% de paulistanos usando mais de dois tipos de transporte.

5 Um novo medicamento para cicatrização está sendo testado e um experimento é feito para estudar o tempo (em dias) de completo fechamento em cortes provenientes de cirurgia. Uma amostra em trinta cobaias forneceu os valores:

```
> dias <- c(15, 17, 16, 15, 17, 14, 17, 16, 16, 17, 15, 18, 14,
+ 17, 15, 14, 15, 16, 17, 18, 18, 17, 15, 16, 14, 18, 18, 16,
+ 15, 14)

> dias.c <- factor(dias, labels = c("14", "15", "16", "17", "18"))
> tab.abs <- table(dias.c)
> total <- sum(tab.abs)
> tab.d <- round(prop.table(table(dias.c)), 2)
> tab.d

dias.c
 14  15  16  17  18
0.17 0.23 0.20 0.23 0.17

> fac.1 <- tab.d[1]
> fac.2 <- fac.1 + tab.d[2]
```

```

> fac.3 <- fac.2 + tab.d[3]
> fac.4 <- fac.3 + tab.d[4]
> fac.5 <- fac.4 + tab.d[5]

```

(a) Organize uma tabela de frequência

Cicatrização	15	17	16	15	17	Total
frequência Abs.	5	7	6	7	5	30
frequência Rel.	0.17	0.23	0.2	0.23	0.17	1.00
frequência Ac.	0.17	0.4	0.6	0.83	1	—

(b) Que porcentagem das observações estão abaixo de 16 dias?

Pela tabela vemos que é 40%.

(c) Classifique como rápida as cicatrizações iguais ou inferiores a 15 dias e como lenta as demais. Faça um diagrama circular indicando as porcentagens para cada classificação.

```

> rapida <- dias <= 15
> lenta <- dias > 15
> dias.r <- factor(rapida, labels = c("lenta", "rapida"))
> table(dias.r)

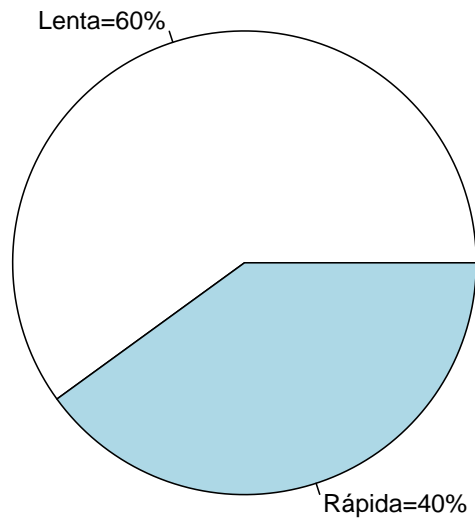
dias.r
lenta rapida
  18     12

> tab.r <- round(prop.table(table(dias.r)), 2)
> tab.r

dias.r
lenta rapida
  0.6     0.4

> pie(tab.r, c("Lenta=60%", "Rápida=40%"))

```



7 Um questionário foi aplicado aos dez funcionários do setor de contabilidade de uma empresa fornecendo os dados apresentados na tabela.

Funcionário	Curso(completo)	Idade	salário(R\$)	Anos de Empresa
1	superior	34	1100,00	5
2	superior	43	1450,00	8
3	médio	31	960,00	6
4	médio	37	960,00	8
5	médio	24	600,00	3
6	médio	25	600,00	2
7	médio	27	600,00	5
8	médio	22	450,00	2
9	fundamental	21	450,00	3
10	fundamental	26	450,00	3

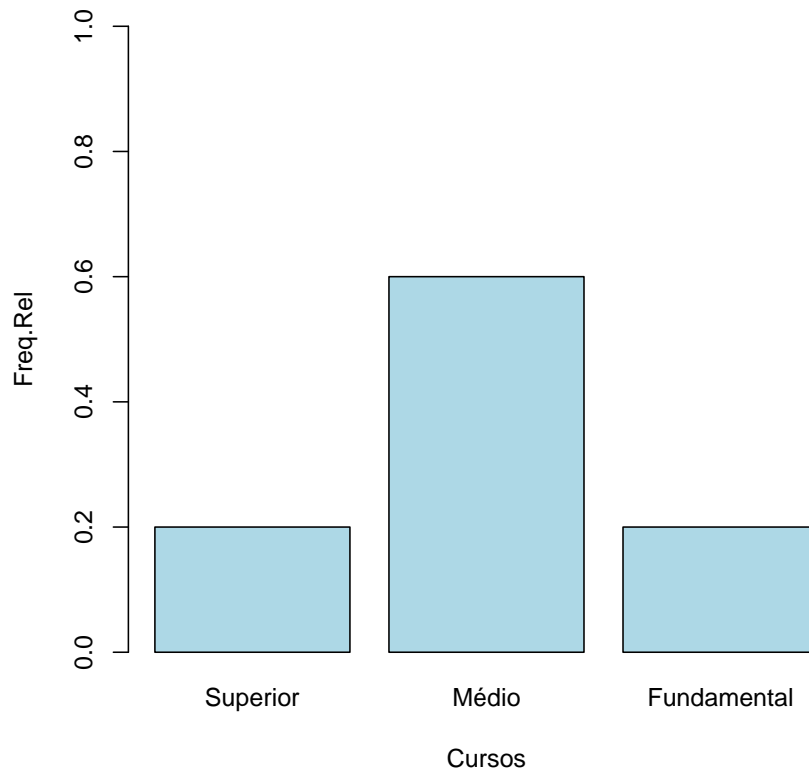
- (a) Classifique cada uma das variáveis.
 variável Curso - Qualitativa Ordinal
 variável Idade - Quantitativa Discreta
 variável salário - Quantitativa contínua
 variável Anos de Empresa - Quantitativa Discreta

(b) Faça uma representação gráfica para a variável Curso.

```
> curso <- rep(c("superior", "médio", "fundamental"), c(2, 6,
+ 2))
> curso.t <- factor(curso, labels = c("Superior", "Médio", "Fundamental"))
> table(curso.t)

curso.t
  Superior      Médio Fundamental
         2         6           2

> tab.curso <- round(prop.table(table(curso.t)), 2)
> barplot(tab.curso, col = c("lightblue"), xlab = "Cursos", ylim = c(0,
+ 1), ylab = "Freq.Rel")
```

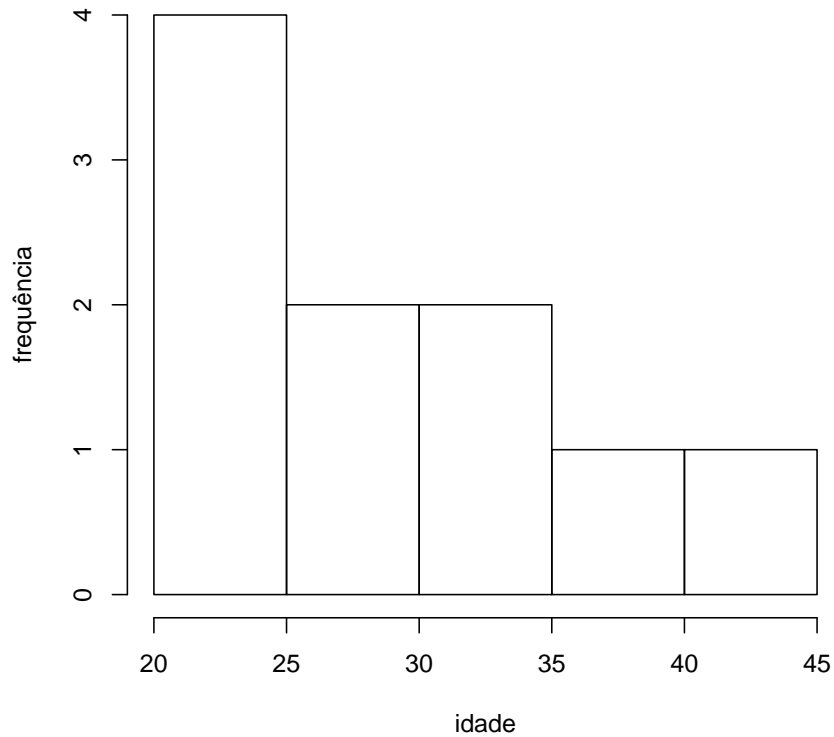


(c) Discuta a melhor forma de construir a tabela de frequência para a variável idade. Construa uma representação gráfica.

Pode-se fazer uma tabela de frequência dividindo a idade em classes de amplitude 5 tendo como o limite inferior da primeira o valor 20 e limite superior da última classe o valor 45.

```
> idade <- c(34, 43, 31, 37, 24, 25, 27, 22, 21, 26)
> hist(idade, main = "Histograma para a Idade", ylab = "frequência")
```

Histograma para a Idade



(d) Repita o item (c) para a variável salário

```
> sal <- c(1100, 1450, 960, 960, 600, 600, 600, 450, 450, 450)
> m <- sal <= 600
> summary(m)
  Mode FALSE  TRUE
logical   4    6

> e <- sal >= 800 & sal <= 1200
> summary(e)
  Mode FALSE  TRUE
logical   7    3

> u <- sal >= 1400
> summary(u)
  Mode FALSE  TRUE
logical   9    1

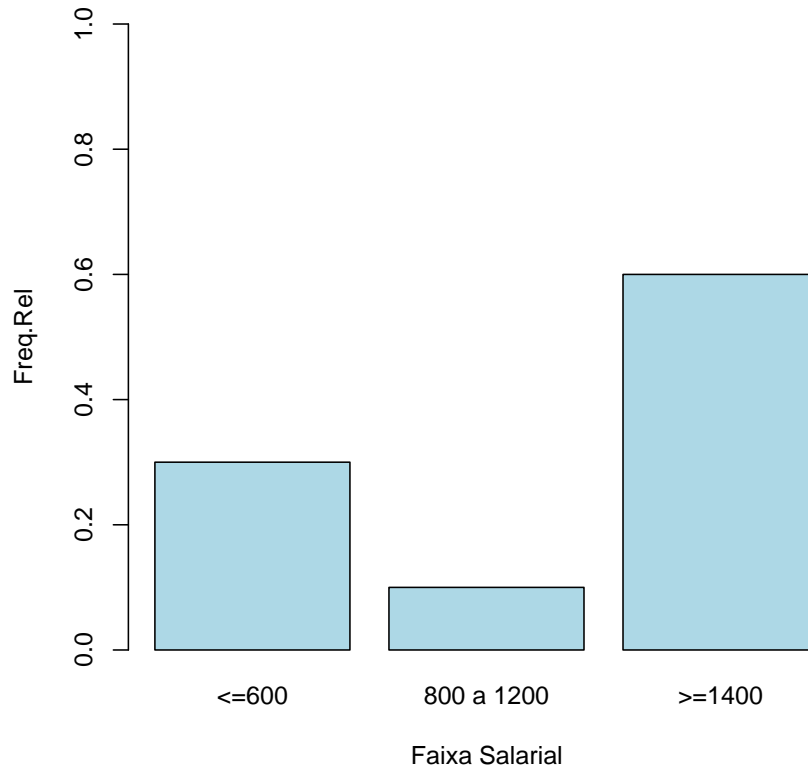
> salario <- rep(c("sal<=600", "800>=sal<=1200", "sal>=1400"),
+               c(6, 3, 1))
> sal.t <- factor(salario, labels = c("<=600", "800 a 1200", ">=1400"))
> table(sal.t)

sal.t
  <=600 800 a 1200  >=1400
      3         1         6
```

```

> tab.sal <- round(prop.table(table(sal.t)), 2)
> barplot(tab.sal, col = c("lightblue"), xlab = "Faixa Salarial",
+         ylim = c(0, 1), ylab = "Freq.Rel")

```



- (e) Considerando apenas os funcionários com mais de três anos de casa, descreva o comportamento da variável salário.

```

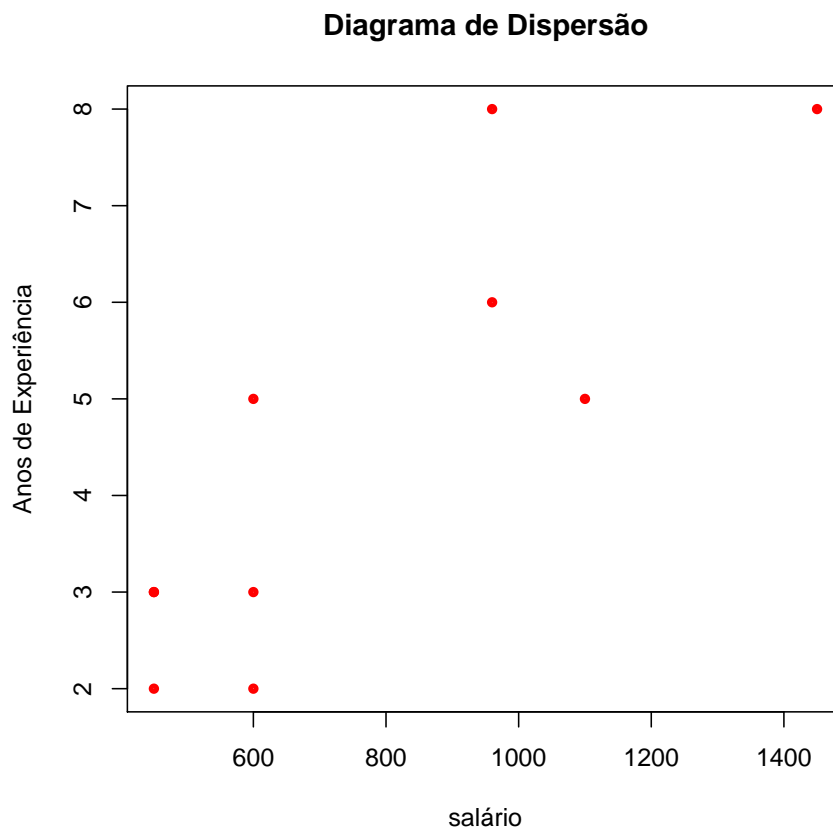
> A.E <- c(5, 8, 6, 8, 3, 2, 5, 2, 3, 3)
> t <- data.frame(sal, A.E)
> t
   sal A.E
1 1100  5
2 1450  8
3  960  6
4  960  8
5  600  3
6  600  2
7  600  5
8  450  2
9  450  3
10 450  3

> summary(sal[A.E > 3])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   600    960    960   1014   1100   1450

```



```
> plot(t, xlab = "salário", ylab = "Anos de Experiência", main = "Diagrama de Dispersão"
+      lwd = 3, cex = 0.5, col = "red")
```



9 Alunos da Escola de Educação Física foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiram correr sem parar. Os dados estão apresentados a seguir.

Faixas	frequência
0 † 4	438
4 † 8	206
8 † 12	125
12 † 16	22
16 † 20	9

- (a) Qual é a variável em estudo?
número de quilômetros percorridos sem parar.
- (b) Construa o histograma.

```
> f.um <- seq(0, 3, l = 438)
> f.dois <- seq(4, 7, l = 206)
> f.três <- seq(8, 11, l = 125)
```

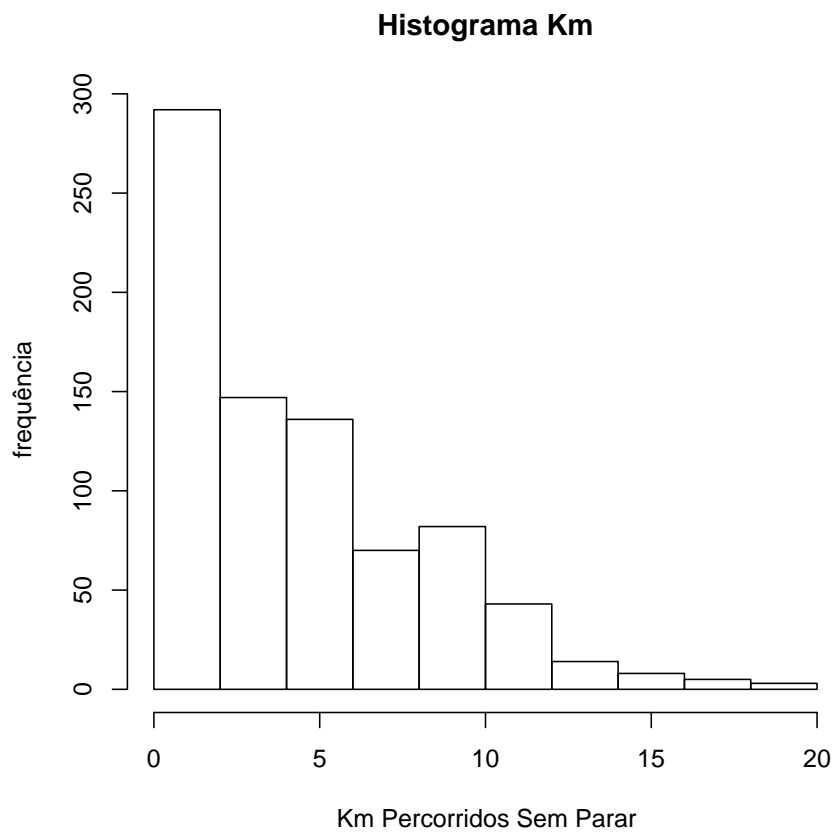
```

> f.quatro <- seq(12, 15, l = 22)
> f.cinco <- seq(16, 19, l = 9)
> dados <- c(f.um, f.dois, f.três, f.quatro, f.cinco)
> head(dados)

[1] 0.000000000 0.006864989 0.013729977 0.020594966 0.027459954 0.034324943
> length(dados)

[1] 800
> hist(dados, main = "Histograma Km", xlab = "Km Percorridos Sem Parar",
+      ylab = "frequência")

```



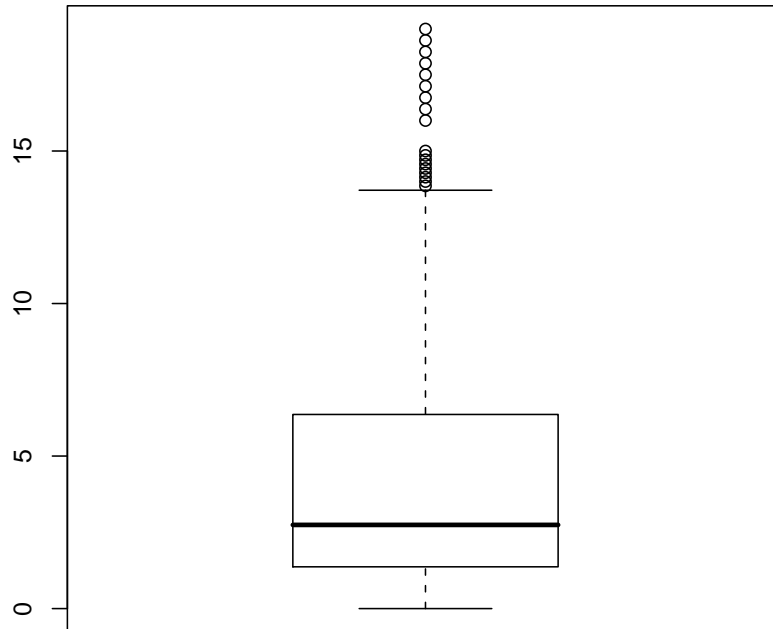
(c) Obtenha o BoxPlot.

```

> boxplot(dados, main = "BoxPlot Km")

```

BoxPlot Km



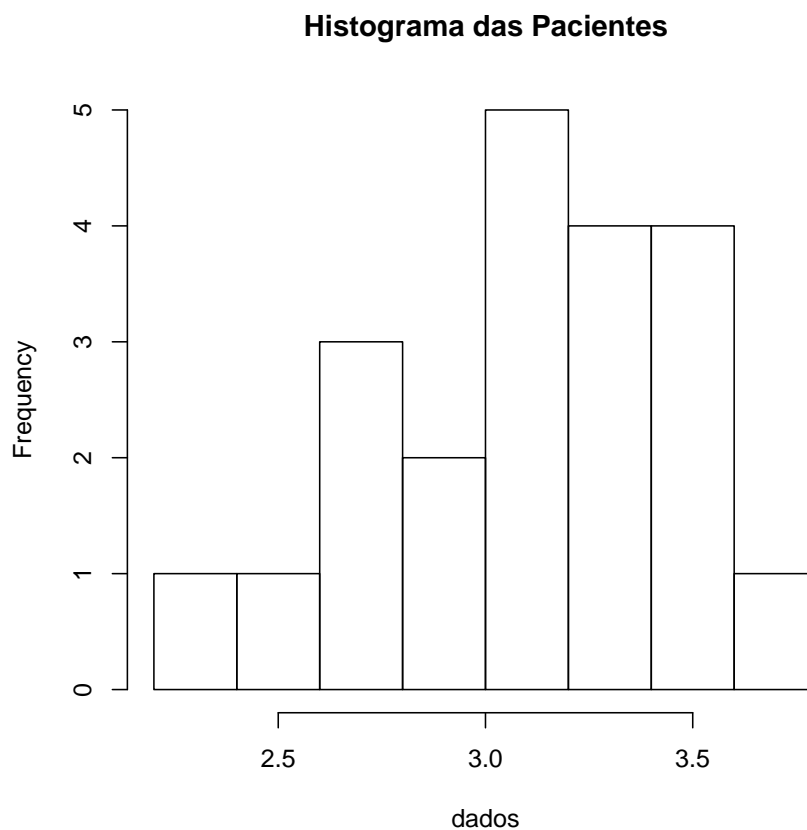
11 Vinte e uma pacientes de uma clínica médica tiveram o seu nível de potássio no plasma medido. Os resultados foram os seguintes:

Nível	frequência
2,25 † 2,55	1
2,55 † 2,75	3
2,75 † 2,95	2
2,95 † 3,15	4
3,15 † 3,35	5
3,35 † 3,65	6

(a) Construa o Histograma.

```
> f.um <- seq(2.25, 2.54, l = 1)
> f.dois <- seq(2.55, 2.74, l = 3)
> f.três <- seq(2.75, 2.94, l = 2)
> f.quatro <- seq(2.95, 3.14, l = 4)
> f.cinco <- seq(3.15, 3.34, l = 5)
> f.seis <- seq(3.35, 3.65, l = 6)
> dados <- c(f.um, f.dois, f.três, f.quatro, f.cinco, f.seis)
> head(dados)
```

```
[1] 2.250 2.550 2.645 2.740 2.750 2.940
> length(dados)
[1] 21
> hist(dados, main = "Histograma das Pacientes")
```



(b) Determine os 1^o, 2^o e 3^o quartis.

```
> summary(dados)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.250  2.940   3.150   3.109  3.350   3.650
```

(c) Qual a porcentagem dos valores que estão acima do nível três?

```
> 100 * (table(dados > 3)/length(dados))
   FALSE    TRUE
 33.33333 66.66667
```

Aproximadamente 66,7% dos valores.

13 O valor médio de comercialização da saca de milho de 60 quilos na Bolsa de Cereais é apresentado a seguir, em reais, para últimos 40 meses.

6.1	6.2	6.7	6.5	6.9	6.3	7.4	7.6	7.7	7.6
7.3	7.7	7.6	7.4	7.2	7.2	7.3	7.6	7.5	7.4
7.5	7.7	8.2	8.3	8.1	8.1	8.1	7.9	7.8	7.4
7.5	7.6	7.5	7.6	7.4	7.3	7.4	7.5	7.5	7.4

(a) Organize os dados em faixas de tamanho 0,4 a partir de 6.

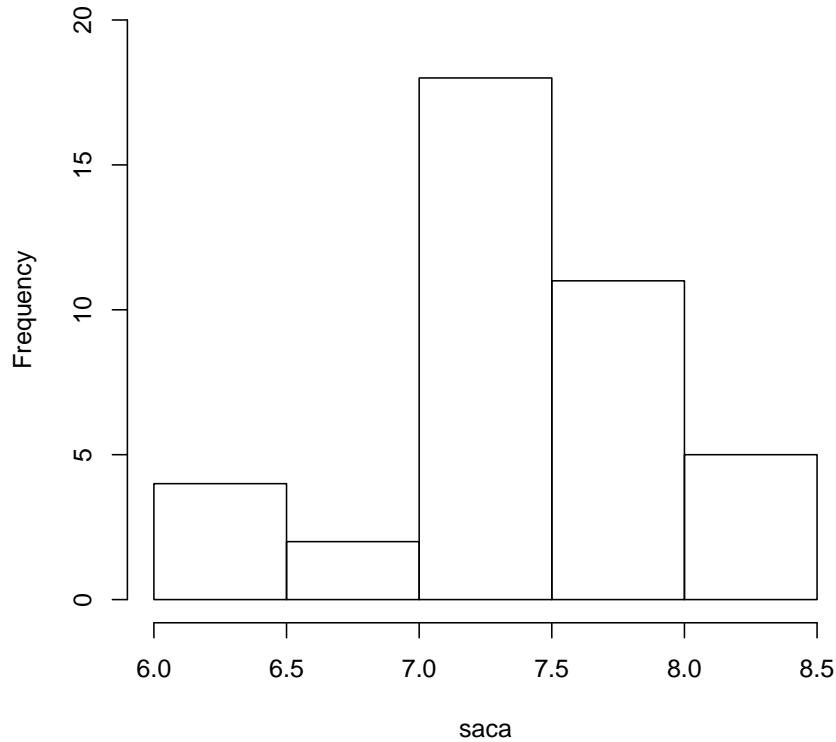
```
> sort(saca)
[1] 6.1 6.2 6.3 6.5 6.7 6.9 7.2 7.2 7.3 7.3 7.3 7.4 7.4 7.4 7.4 7.4 7.4 7.4 7.5
[20] 7.5 7.5 7.5 7.5 7.5 7.6 7.6 7.6 7.6 7.6 7.6 7.7 7.7 7.7 7.8 7.9 8.1 8.1 8.1
[39] 8.2 8.3
> f1 <- saca < 6.4
> f2 <- saca >= 6.4 & saca < 6.8
> f3 <- saca >= 6.8 & saca < 7.2
> f4 <- saca >= 7.2 & saca < 7.6
> f5 <- saca >= 7.6 & saca < 8
> f6 <- saca >= 8 & saca < 8.4
> tab1 <- round(prop.table(table(f1)), 3)
> tab2 <- round(prop.table(table(f2)), 3)
> tab3 <- round(prop.table(table(f3)), 3)
> tab4 <- round(prop.table(table(f4)), 3)
> tab5 <- round(prop.table(table(f5)), 3)
> tab6 <- round(prop.table(table(f6)), 3)
> F1 <- tab1[2]
> F2 <- F1 + tab2[2]
> F3 <- F2 + tab3[2]
> F4 <- F3 + tab4[2]
> F5 <- F4 + tab5[2]
> F6 <- F5 + tab6[2]
```

Valor	n_i	f_i	F_i
6 \vdash 6,4	3	0.075	0.075
6,4 \vdash 6,8	2	0.05	0.125
6,8 \vdash 7,2	1	0.025	0.15
7,2 \vdash 7,6	18	0.45	0.6
7,6 \vdash 8,0	11	0.275	0.875
8,0 \vdash 8,4	5	0.125	1

(b) Construa o Histograma

```
> hist(saca, main = "histograma Mensal da Saca de Milho", ylim = c(0,
+ 20))
```

histograma Mensal da Saca de Milho



(c) Calcule o segundo Quartil

```
> q <- summary(saca)
> q.2 <- q[3]
> q.2
Median
 7.5
```

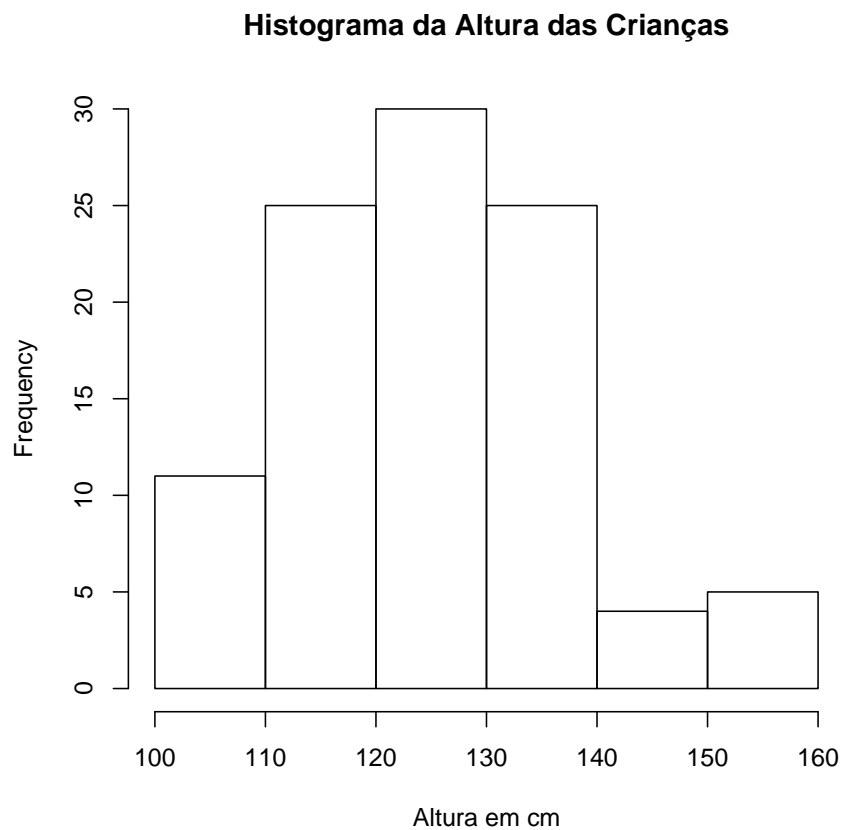
15 A tabela a seguir apresenta as frequências relativas de ocorrências de faixas de altura (em cm) para uma amostra de 100 crianças de 12 anos de idade.

Faixas	Frequência Relativa
100 - 110	0,10
110 - 120	0,25
120 - 130	0,30
130 - 140	0,25
140 - 160	0,10

(a) Construa o Histograma

```
> f.um <- seq(100, 109, l = 10)
> f.dois <- seq(110, 119, l = 25)
> f.três <- seq(120, 129, l = 30)
> f.quatro <- seq(130, 139, l = 25)
```

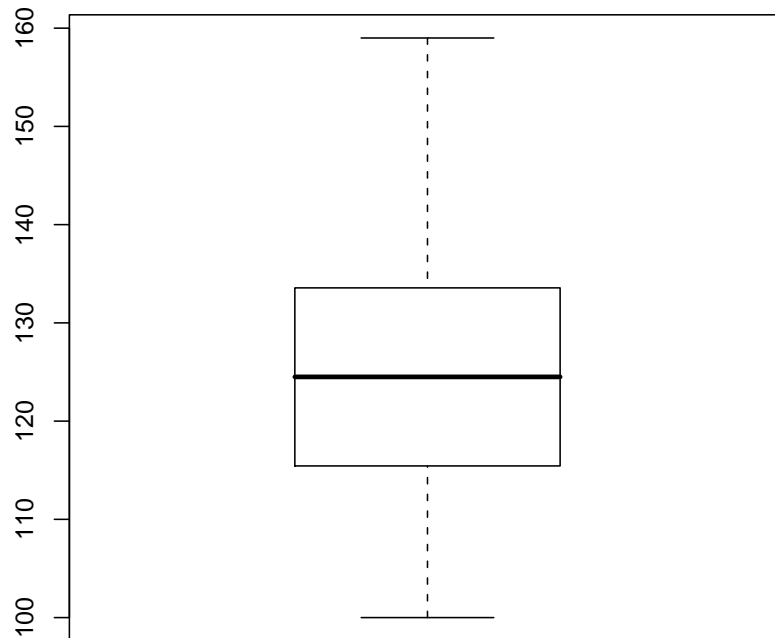
```
> f.cinco <- seq(140, 159, l = 10)
> altura <- c(f.um, f.dois, f.três, f.quatro, f.cinco)
> head(altura)
[1] 100 101 102 103 104 105
> length(altura)
[1] 100
> hist(altura, main = "Histograma da Altura das Crianças", xlab = "Altura em cm")
```



(b) Construa o Box Plot

```
> boxplot(altura, main = "BoxPlot da Altura das Crianças")
```

BoxPlot da Altura das Crianças



(c) Desejando-se separar os 15% mais altos qual seria o ponto de corte?

```
> d <- sort(altura)
> corte <- d[c(86, 100)]
> corte <- min(corte)
> corte
[1] 137.5
```

17 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 3,5Kg em um mês. A seguir, apresentamos os dados referentes ao ganho, em quilos, para essa nova ração, aplicada durante um mês em 200 animais nas condições acima.

Ganho	Frequência
1 † 2	45
2 † 3	83
3 † 4	52
4 † 5	15
5 † 6	4
6 † 7	1

(a) Construa o Histograma


```

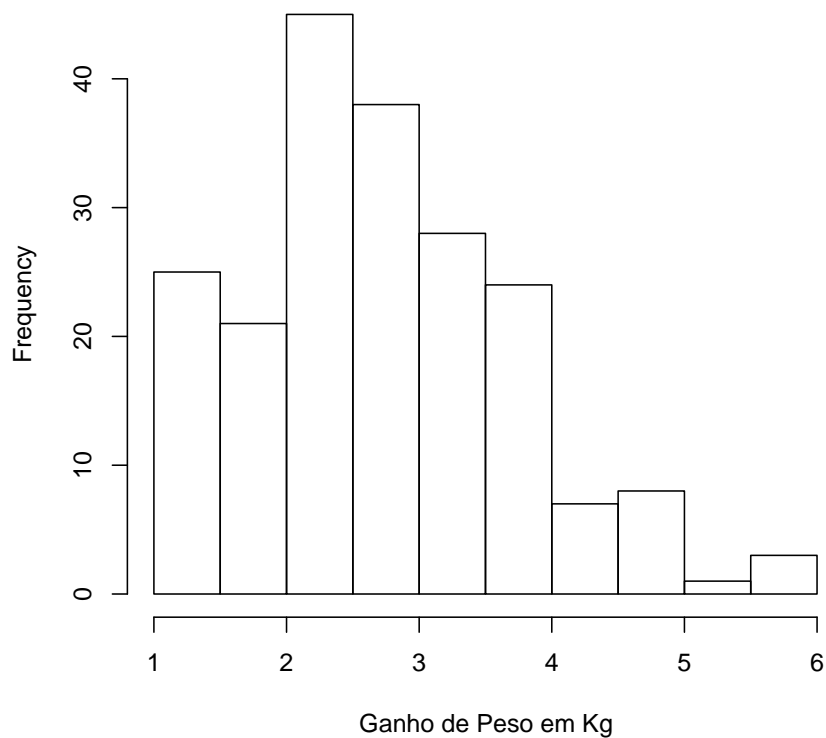
> f.um <- seq(1, 1.9, l = 45)
> f.dois <- seq(2, 2.9, l = 83)
> f.tres <- seq(3, 3.9, l = 52)
> f.quatro <- seq(4, 4.9, l = 15)
> f.cinco <- seq(5, 5.9, l = 4)
> f.seis <- seq(6, 6.9, l = 1)
> rac <- c(f.um, f.dois, f.tres, f.quatro, f.cinco, f.seis)
> head(rac)

[1] 1.000000 1.020455 1.040909 1.061364 1.081818 1.102273
> length(rac)

[1] 200
> hist(rac, main = "Histograma do Ganho de Peso", xlab = "Ganho de Peso em Kg")

```

Histograma do Ganho de Peso



(b) Determine o 1^o, 2^o e 3^o quartis.

```

> s <- summary(rac)
> q.1 <- s[2]
> q.2 <- s[3]
> q.3 <- s[5]
> res <- c(q.1, q.2, q.3)
> res

```

```

1st Qu.  Median 3rd Qu.
  2.052   2.598   3.375

```

- (c) Você acha que a nova ração é mais eficiente que a tradicional? Justifique.
Não, basta verificarmos que a média é menor que 3,5 Kg e o 3^o quartil é menor que este valor também, ou seja 75% dos suínos tiveram ganho abaixo que o valor de referência.

19 Como parte de uma avaliação médica em uma certa universidade, foi medida a frequência cardíaca dos alunos do primeiro ano. Os dados são apresentados em seguida.

Frequência Cardíaca	frequência
60 † 65	11
65 † 70	35
70 † 75	68
75 † 80	20
80 † 85	12
85 † 90	10
90 † 95	1
95 † 100	3

- (a) Obtenha o Histograma

```
> f.um <- seq(60, 64, l = 11)
> f.dois <- seq(65, 69, l = 35)
> f.tres <- seq(70, 74, l = 68)
> f.quatro <- seq(75, 79, l = 20)
> f.cinco <- seq(80, 84, l = 12)
> f.seis <- seq(85, 89, l = 10)
> f.sete <- seq(90, 94, l = 1)
> f.oito <- seq(95, 100, l = 3)
> f.card <- c(f.um, f.dois, f.tres, f.quatro, f.cinco, f.seis,
+   f.sete, f.oito)
> head(f.card)

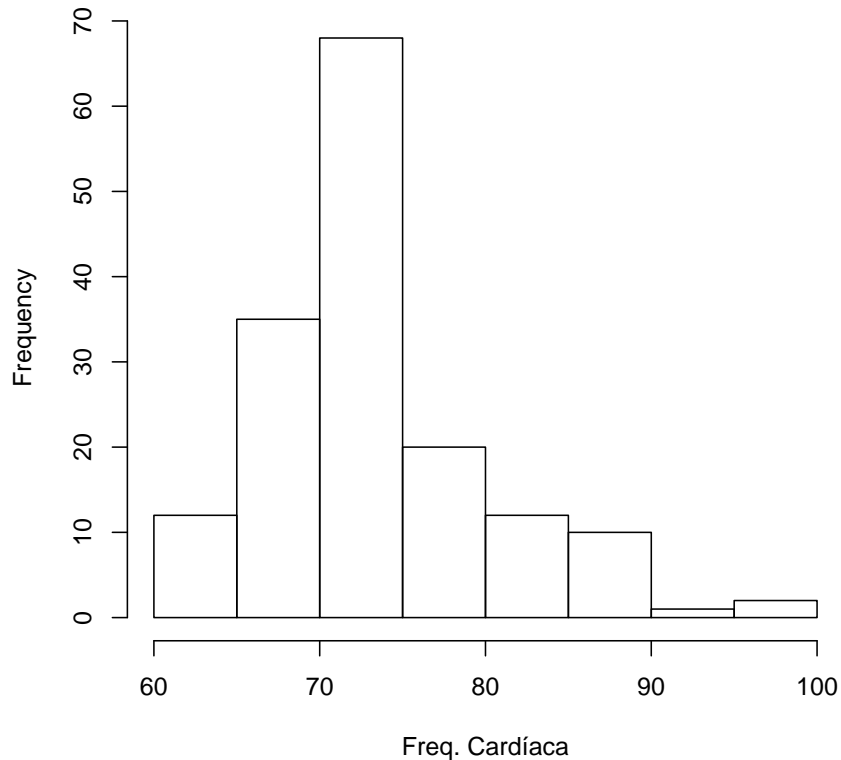
[1] 60.0 60.4 60.8 61.2 61.6 62.0

> length(f.card)

[1] 160

> hist(f.card, main = "Histograma da frequência Cardíaca dos Alunos",
+   xlab = "Freq. Cardíaca")
```

Histograma da frequência Cardíaca dos Alunos



- (b) Frequências cardíacas que estejam abaixo de 62 ou acima de 92 requerem acompanhamento médico. Qual é a porcentagem de alunos nessas condições?

```
> min <- 100 * (table(f.card < 62)/length(f.card))
> max <- 100 * (table(f.card > 92)/length(f.card))
> res <- max + min
> res
```

```
FALSE TRUE
 195    5
```

- (c) Uma frequência ao redor de 72 batidas por minuto é considerada padrão. Você acha que de modo geral esses alunos se encaixam nesse caso?

```
> summary(f.card)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 60.00  68.38   72.00   73.12  76.11  100.00
```

Sim, basta olharmos a mediana e valor médio que não se diferenciam muito.

- 21 Vinte baterias para automóveis de uma certa marca foram testadas quanto à sua vida útil. O teste simula a utilização da bateria, acelerando seu desgaste de modo a criar uma réplica da situação real. Os resultados da durabilidade (em meses) são apresentados a seguir.

- (a) Construa o Histograma

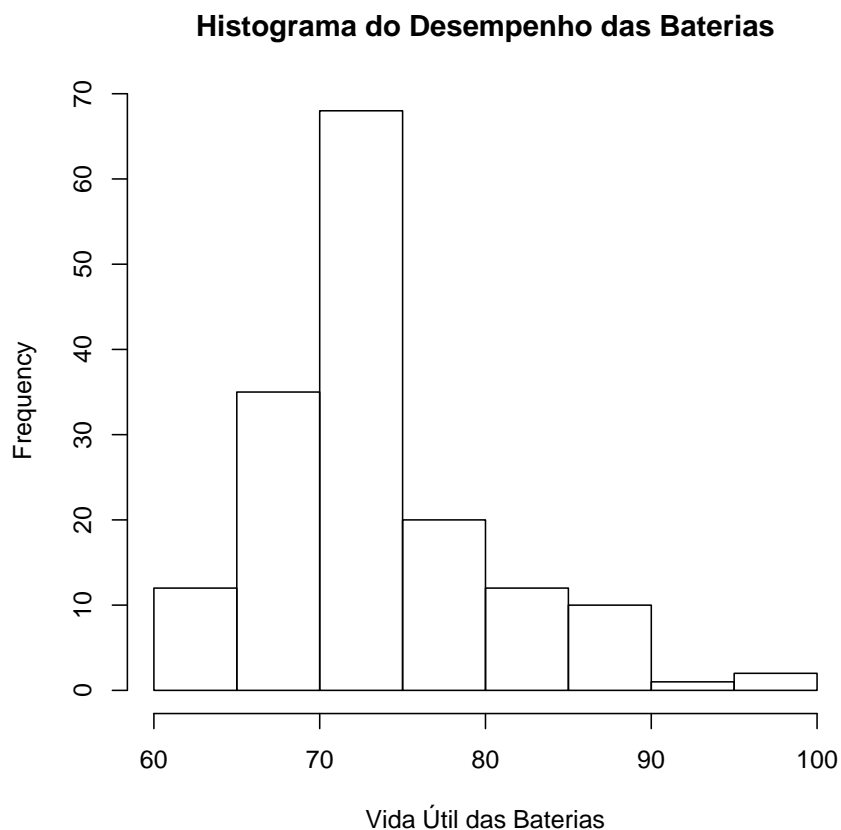
```
> f.um <- seq(0, 2, l = 20)
> f.dois <- seq(3, 5, l = 50)
```

Durabilidade	Frequência Relativa
0 † 3	0,02
3 † 6	0,05
6 † 9	0,15
9 † 12	0,25
12 † 15	0,30
15 † 20	0,23

```

> f.tres <- seq(6, 8, l = 150)
> f.quatro <- seq(9, 11, l = 250)
> f.cinco <- seq(12, 14, l = 300)
> f.seis <- seq(15, 20, l = 230)
> f.bat <- c(f.um, f.dois, f.tres, f.quatro, f.cinco, f.seis)
> head(f.bat)
[1] 0.0000000 0.1052632 0.2105263 0.3157895 0.4210526 0.5263158
> length(f.bat)
[1] 1000
> hist(f.card, main = "Histograma do Desempenho das Baterias",
+     xlab = "Vida Útil das Baterias")

```



- (b) Se a amostra for considerada representativa do desempenho dessa marca de bateria, quantas, em 1000 fabricadas, serão repostas pelo fabricante, se ele oferece 6 meses de garantia?

```
> seis.meses <- f.bat < 6
> summary(seis.meses)
```

```
   Mode  FALSE   TRUE
logical   930    70
```

Pela saída do R associado ao valor TRUE, 70 baterias.

23 Os dados a seguir representam indivíduos que foram contaminados pelo veneno de um certo tipo de inseto e submetidos a tratamento (três diferentes tipos).

(a) Através do computador crie uma planilha com os dados apresentados. Baseando-se nesta planilha:

i. Classifique cada uma das variáveis.

- Idade, Variável quantitativa discreta;
- Diag, Variável quantitativa contínua;
- Recup, Variável quantitativa contínua;
- Tratam, Variável qualitativa ordinal;
- Coag, Variável qualitativa nominal.

ii. Construa uma tabela de frequência (através do computador) para a variável Diag. Utilize 5 classes.

```
> Num <- c(19, 4, 27, 7, 14, 5, 11, 10, 25, 6, 16, 20, 13, 15,
+ 8, 18, 12, 24, 21, 22, 3, 2, 23, 26, 17, 9)
> Idade <- c(28, 15, 76, 15, 21, 11, 16, 16, 47, 18, 40, 24, 32,
+ 31, 10, 31, 31, 46, 21, 39, 15, 9, 75, 54, 35, 18)
> Diag <- c(7, 52, 30, 53, 3, 46, 55, 54, 13, 59, 20, 3, 9, 9,
+ 44, 9, 10, 13, 1, 17, 53, 42, 30, 18, 12, 58)
> Recup <- c(3, 45, 23, 46, 2, 42, 47, 47, 12, 51, 11, 1, 3, 3,
+ 40, 3, 4, 11, 2, 8, 46, 39, 22, 16, 5, 50)
> Coag <- c("nao", "nao", "sim", "sim", "nao", "nao", "nao", "sim",
+ "sim", "nao", "sim", "nao", "nao", "nao", "sim", "sim", "sim",
+ "sim", "sim", "sim", "sim", "nao", "sim", "nao", "sim", "sim")
> Coag <- as.factor(Coag)
> summary(Coag)
nao sim
 11  15

> Tratam <- c(2, 1, 3, 1, 2, 1, 1, 1, 3, 2, 3, 2, 2, 2, 1, 2, 2,
+ 3, 2, 3, 1, 1, 3, 3, 2, 2)
> Coag <- factor(Coag == "sim", level = c(0, 1))
> dados <- data.frame(Num, Idade, Diag, Recup, Coag, Tratam)
> head(dados)
  Num Idade Diag Recup Coag Tratam
1  19    28    7     3    0      2
2   4    15   52    45    0      1
```

```

3 27 76 30 23 1 3
4 7 15 53 46 1 1
5 14 21 3 2 0 2
6 5 11 46 42 0 1
> attach(dados)
> names(dados)
[1] "Num" "Idade" "Diag" "Recup" "Coag" "Tratam"
> summary(Diag)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00  9.25   19.00   27.69  50.50   59.00
> table(Diag)
Diag
 1  3  7  9 10 12 13 17 18 20 30 42 44 46 52 53 54 55 58 59
 1  2  1  3  1  1  2  1  1  1  2  1  1  1  1  2  1  1  1  1
> tab <- matrix(c("1 a 12", "12 a 24", "24 a 36", "36 a 48", "48 a 60",
+ "Total", 8, 6, 2, 3, 7, 26), ncol = 2)
> dimnames(tab) <- list(c("", "", "", "", "", ""), c("Faixas",
+ "frequencia"))

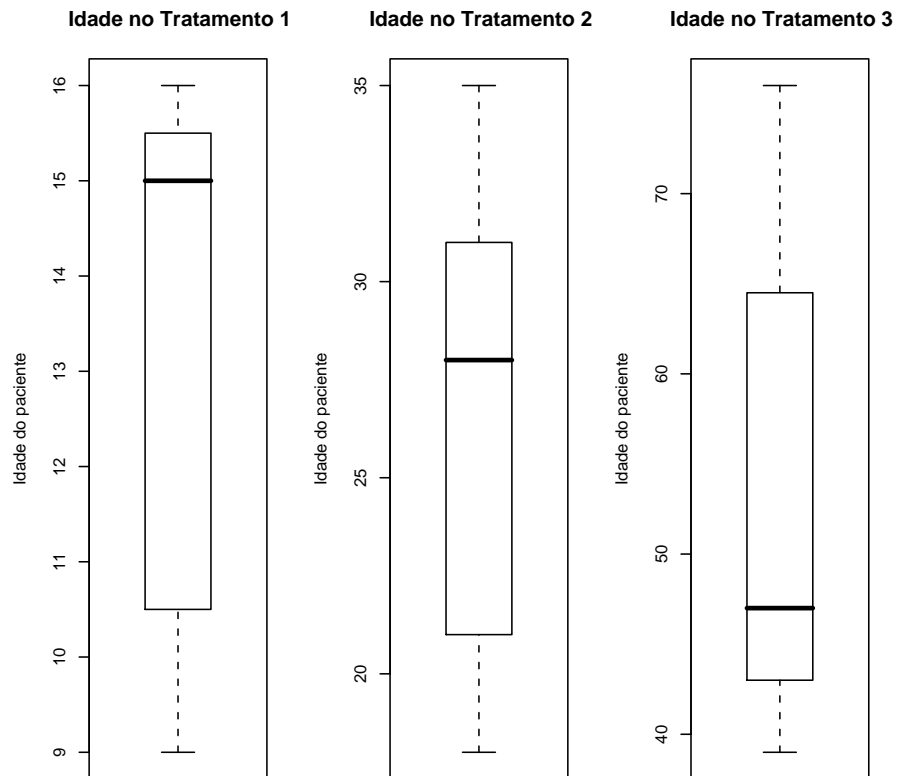
```

- iii. Através de representação gráfica adequada, compare os três tratamentos com relação à idade dos pacientes. Você diria que a idade se distribui homogeneamente nos três tratamentos?

```

> tapply(Idade, Tratam, mean)
      1      2      3
13.37500 26.36364 53.85714
> tapply(Idade, Tratam, median)
 1  2  3
15 28 47
> par(mfrow = c(1, 3))
> boxplot(Idade[Tratam == 1], main = "Idade no Tratamento 1", ylab = "Idade do paciente")
> boxplot(Idade[Tratam == 2], main = "Idade no Tratamento 2", ylab = "Idade do paciente")
> boxplot(Idade[Tratam == 3], main = "Idade no Tratamento 3", ylab = "Idade do paciente")

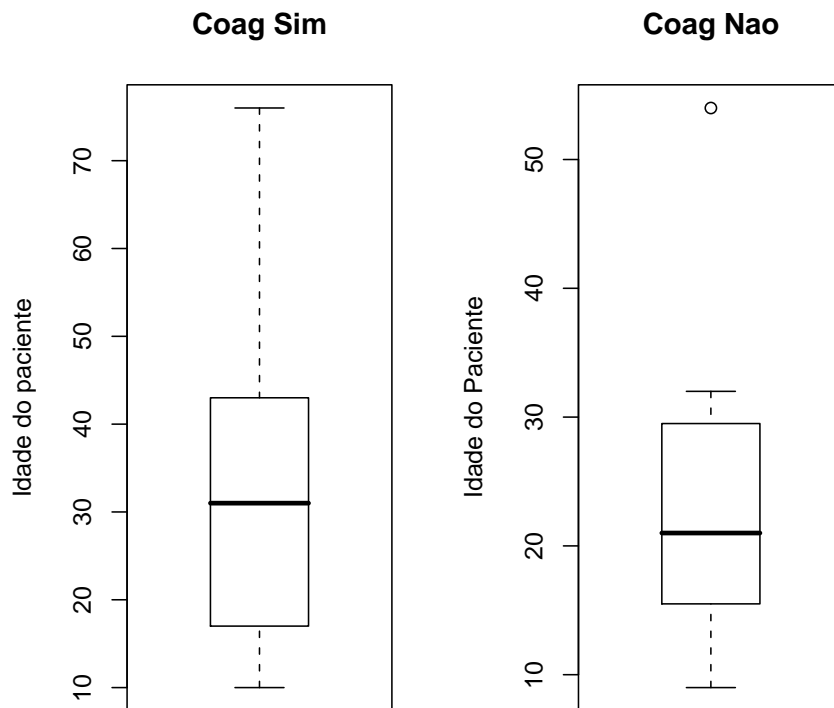
```



Não, a idade não se distribui homogeneamente, verificamos isso através das idades médias e medianas por tratamento e os próprios Boxplots que nos indicam heterogeneidade entre os três tratamentos.

iv. Repita para a variável Coag.

```
> tapply(Idade, Coag, median)
 0  1
21 31
> tapply(Idade, Coag, mean)
      0      1
23.54545 34.33333
> par(mfrow = c(1, 2))
> boxplot(Idade[Coag == 1], main = "Coag Sim", ylab = "Idade do paciente")
> boxplot(Idade[Coag == 0], main = "Coag Nao", ylab = "Idade do Paciente")
```



(b) Suponha que ao invés de trabalhar com a variável Idade, cria-se uma nova variável denominada Etário, assumindo valor 0 se a Idade for menor que 29 anos e 1 caso contrário.

i. Explique com você poderia criar tal variável.

Dividimos a variável Idade em duas categorias, uma com os casos que a Idade for menor que 29 e outra com os maiores que 29 e atribuímos a primeira categoria código zero e a outra categoria código 1 e guardamos os resultados em um objeto chamado Etario, como segue os comandos abaixo.

```
> Etario <- Idade >= 29
> Etario

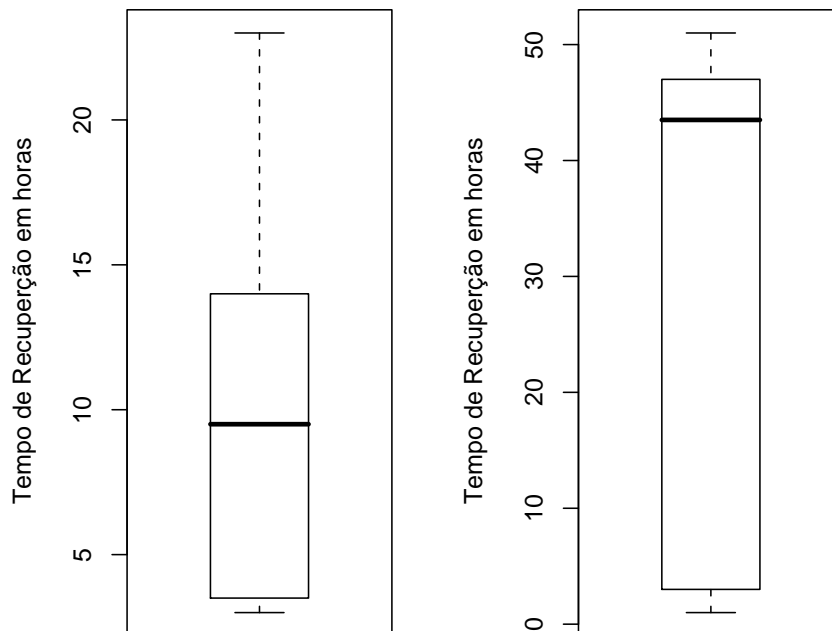
[1] FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE
[13] TRUE TRUE FALSE TRUE TRUE TRUE FALSE TRUE FALSE FALSE TRUE TRUE
[25] TRUE FALSE

> Etario <- as.factor(Etario)
> Etario <- factor(Etario == "TRUE", level = c(0, 1))
```

ii. Construa um BoxPlot para a variável Recup, para cada grupo etário. Com base nos gráficos, você diria que o tempo de recuperação é diferente dependendo da faixa etária? Justifique.

```
> par(mfrow = c(1, 2))
> boxplot(Recup[Etario == 1], main = "Faixa Etária maior que 29 anos",
+       ylab = "Tempo de Recuperação em horas")
> boxplot(Recup[Etario == 0], main = "Faixa Etária menor que 29 anos",
+       ylab = "Tempo de Recuperação em horas")
```

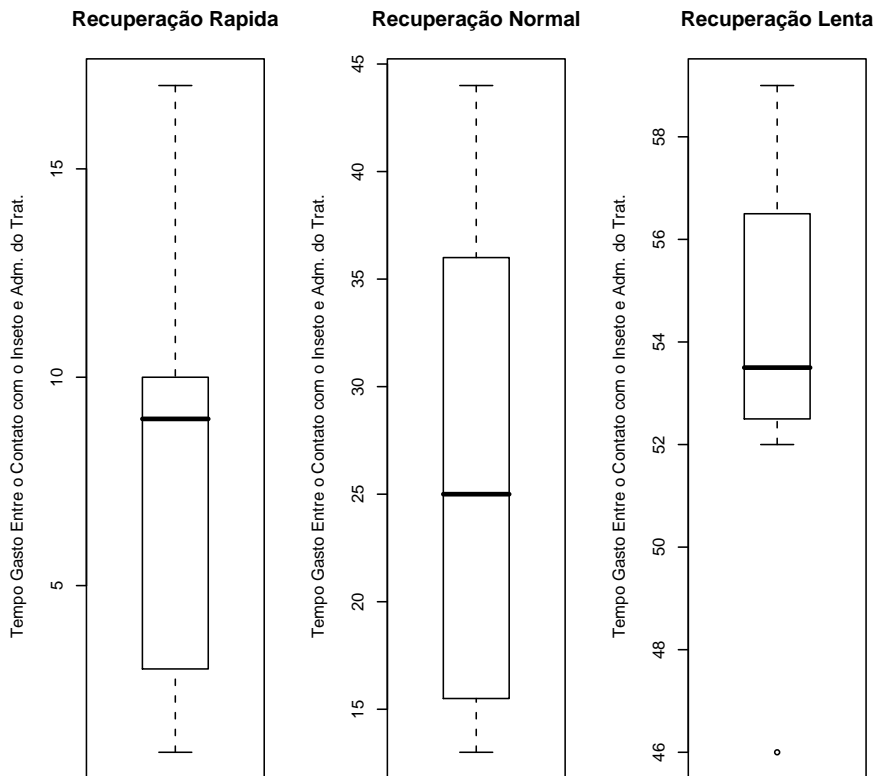

Faixa Etária maior que 29 ano Faixa Etária menor que 29 anc



Sim, quanto a faixa etária maior que 29 nove anos apresenta um tempo maior bem acima que a outra faixa, o que faz sentido já que nesta faixa etária as idades são maiores, isso fica evidente na linha mediana da caixa do BoxPlot.

- (c) Uma nova variável denominada Cura é criada: Cura será rápida se Recup for menor ou igual a 10, será normal entre 10 e 40(inclusive) e será lenta para Recup acima de 40. Verifique, graficamente, se pacientes em cada uma das categorias de Cura apresentam diferenças no que se refere ao tempo entre o contato com o inseto e a administração do tratamento.

```
> rapida <- Recup <= 10
> rapida <- as.factor(rapida)
> rapida <- factor(rapida == "TRUE", level = c(0, 1))
> normal <- Recup > 10 & Recup <= 40
> normal <- as.factor(normal)
> normal <- factor(normal == "TRUE", level = c(0, 1))
> lenta <- Recup > 40
> lenta <- as.factor(lenta)
> lenta <- factor(lenta == "TRUE", level = c(0, 1))
> par(mfrow = c(1, 3))
> boxplot(Diag[rapida == 1], main = "Recuperação Rápida", ylab = "Tempo Gasto Entre o Co
> boxplot(Diag[normal == 1], main = "Recuperação Normal", ylab = "Tempo Gasto Entre o Co
> boxplot(Diag[lenta == 1], main = "Recuperação Lenta", ylab = "Tempo Gasto Entre o Cont
```



25 Um empreendimento imobiliário consistiu da construção de dois edifícios residenciais, com apartamentos de 1 dormitório e área útil de 53 metros quadrados. Em uma primeira, fase foi construído apenas um edifício, denominado Bloco A. Para conclusão do Bloco B houve uma troca de empreiteiras. Suspeita-se que seus apartamentos foram construídos com metragem diferente daquela especificada na escritura.

- (a) Explore o arquivo de dados. Qual o número de apartamentos no empreendimento? Quantos apartamentos existem por prédio? E por andar?

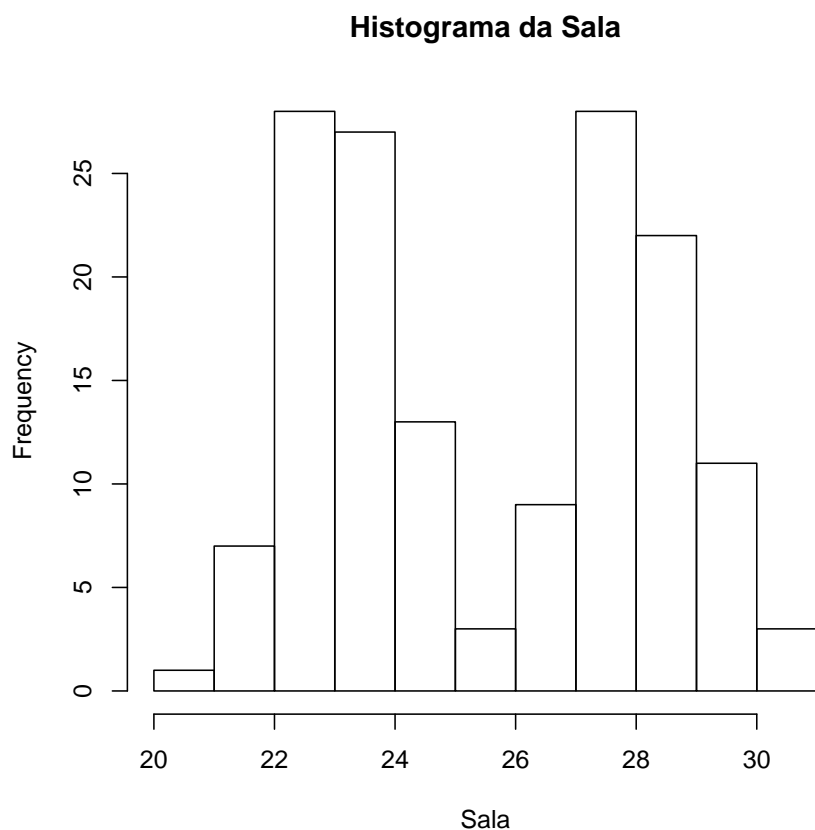
```
> Area <- read.table("http://www.ime.usp.br/~noproest/areas.txt",
+   header = T)
> head(Area)
> attach(Area)
> dim(Area)
[1] 152 10
> names(Area)
 [1] "Id"      "Bloco"   "Andar"   "Final"   "Sala"    "Cozinha"
 [7] "Banheiro" "Dorm"    "Rachadura" "Infiltr"
> sum(Bloco == "A")
[1] 76
> sum(Bloco == "B")
```

```
[1] 76
```

```
> summary(factor(Andar[Bloco == "A"]))  
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19  
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  
> summary(factor(Andar[Bloco == "B"]))  
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19  
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
```

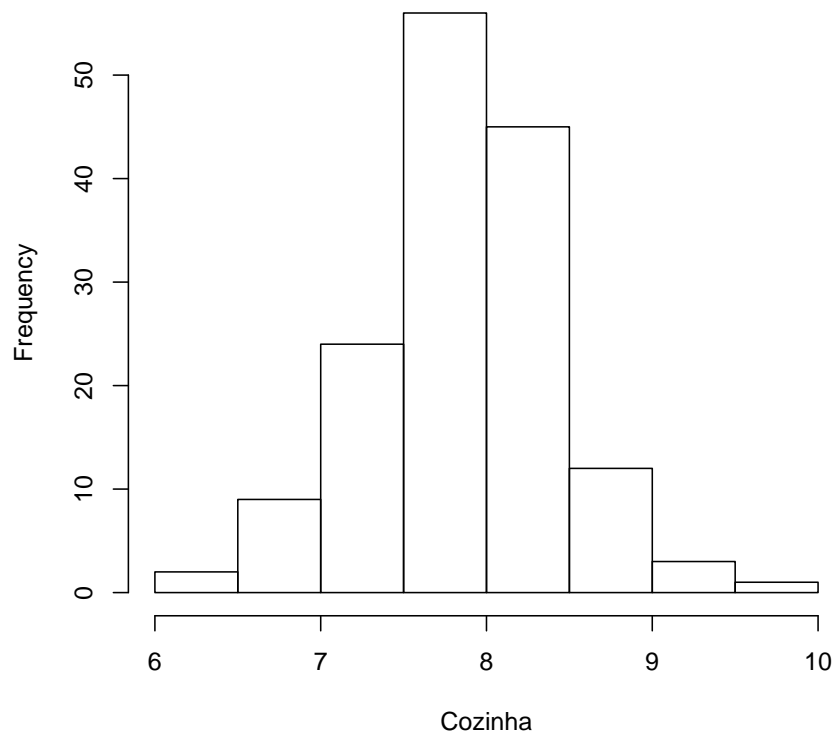
- (b) Construa tabelas de frequência para cada uma das variáveis quantitativas contínuas e faça gráficos adequados.

```
> summary(Sala)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
20.90  23.10  25.30  25.61  27.92  30.90  
> hist(Sala, main = "Histograma da Sala")
```



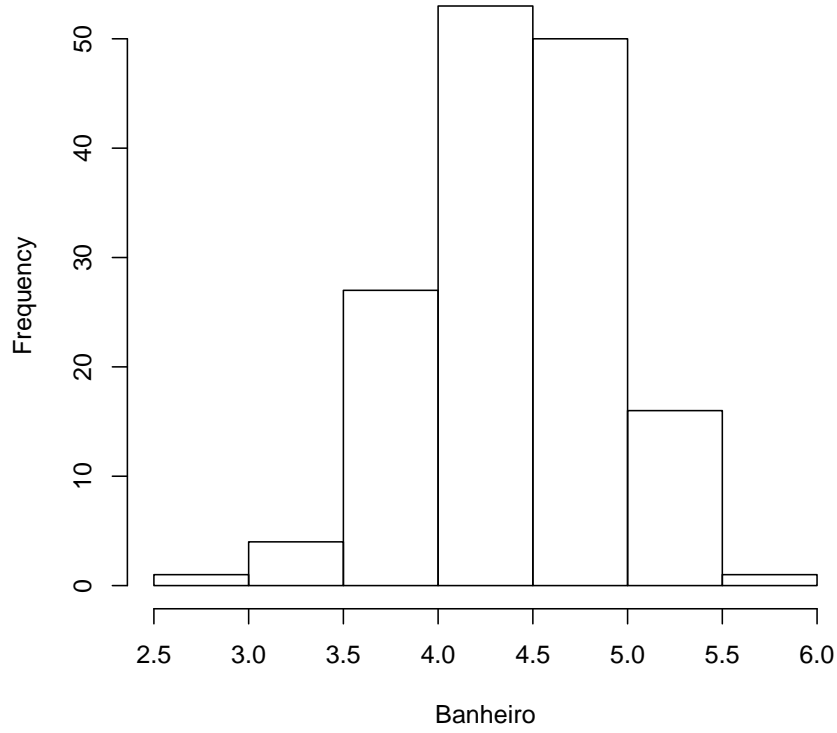
```
> summary(Cozinha)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
6.400  7.600  7.900  7.905  8.225  9.600  
> hist(Cozinha, main = "Histograma da Cozinha")
```

Histograma da Cozinha



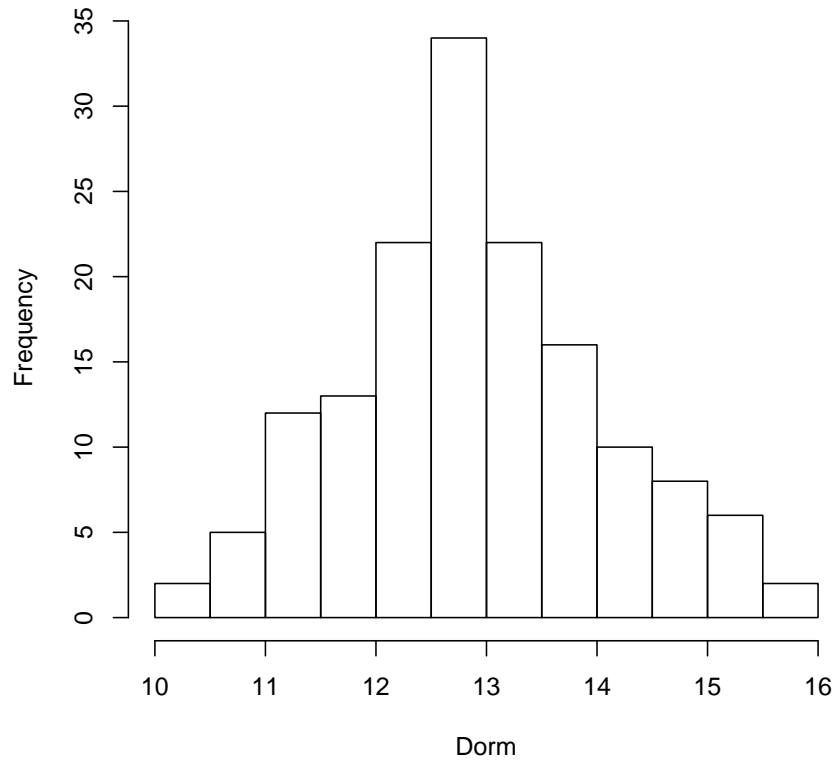
```
> summary(Banheiro)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.900  4.100  4.500  4.457  4.800  5.800
> hist(Banheiro, main = "Histograma do Banheiro")
```

Histograma do Banheiro



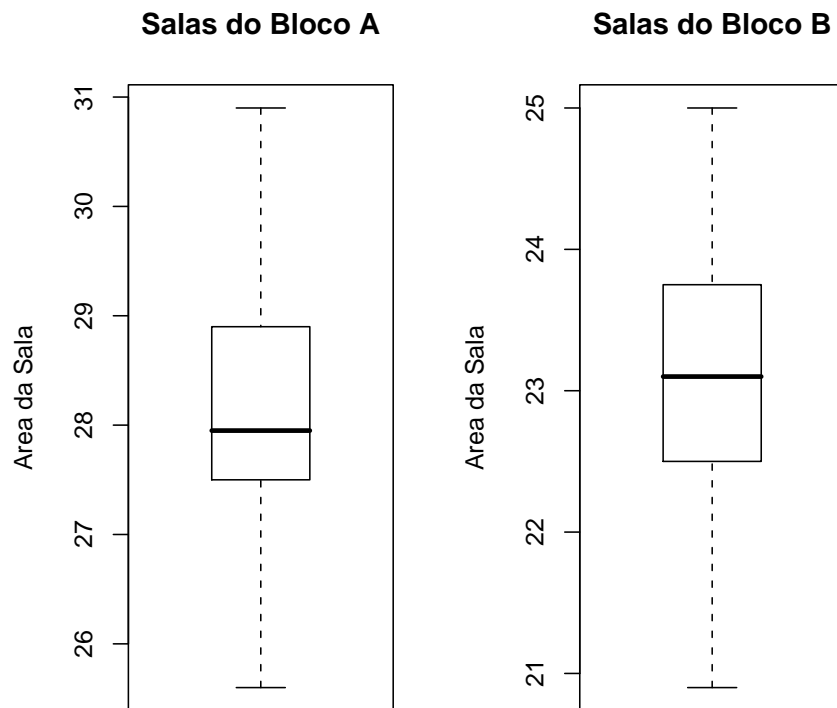
```
> summary(Dorm)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 10.00  12.20  12.95  12.94  13.60  15.90
> hist(Dorm, main = "Histograma do Dormitório")
```

Histograma do Dormitório



- (c) Repita o item (b), para cada bloco, separadamente. Construa gráficos do tipo BoxPlot e compare as áreas para cada cômodo considerado.

```
> par(mfrow = c(1, 2))  
> boxplot(Sala[Bloco == "A"], main = "Salas do Bloco A", ylab = "Area da Sala")  
> boxplot(Sala[Bloco == "B"], main = "Salas do Bloco B", ylab = "Area da Sala")
```

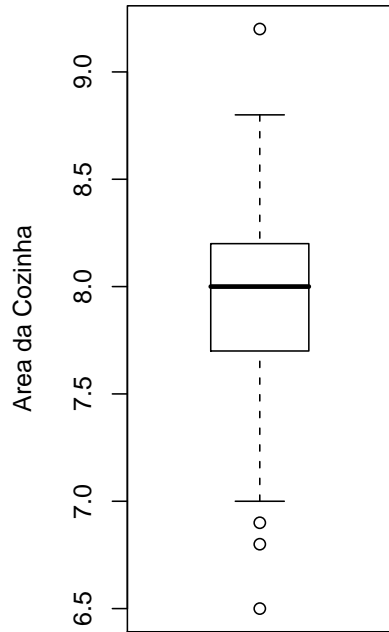


```

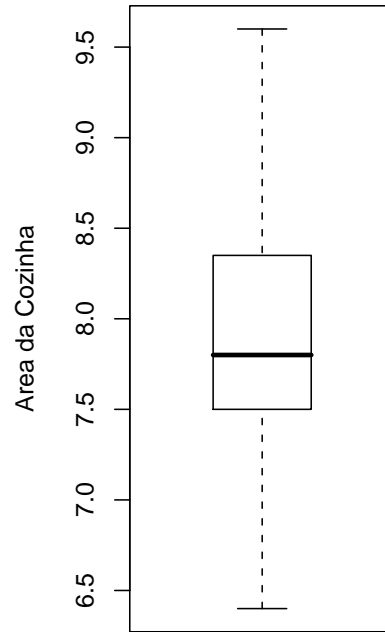
> par(mfrow = c(1, 2))
> boxplot(Cozinha[Bloco == "A"], main = "Cozinhas do Bloco A",
+         ylab = "Area da Cozinha")
> boxplot(Cozinha[Bloco == "B"], main = "Cozinhas do Bloco B",
+         ylab = "Area da Cozinha")

```

Cozinhas do Bloco A

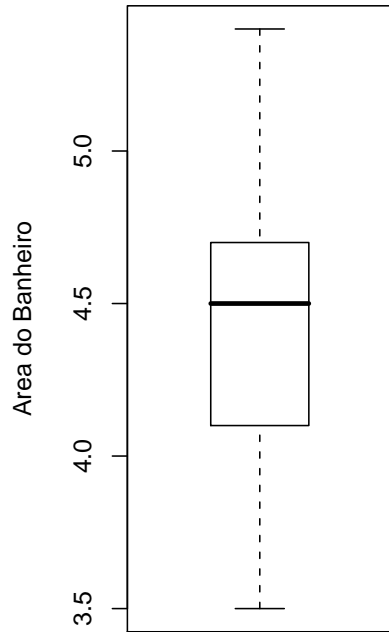


Cozinhas do Bloco B

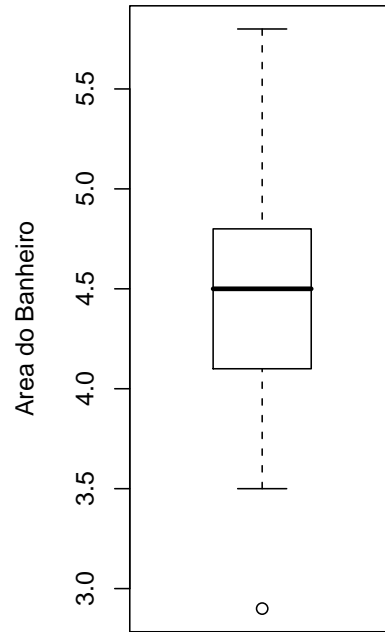


```
> par(mfrow = c(1, 2))  
> boxplot(Banheiro[Bloco == "A"], main = "Banheiros do Bloco A",  
+         ylab = "Area do Banheiro")  
> boxplot(Banheiro[Bloco == "B"], main = "Banheiros do Bloco B",  
+         ylab = "Area do Banheiro")
```

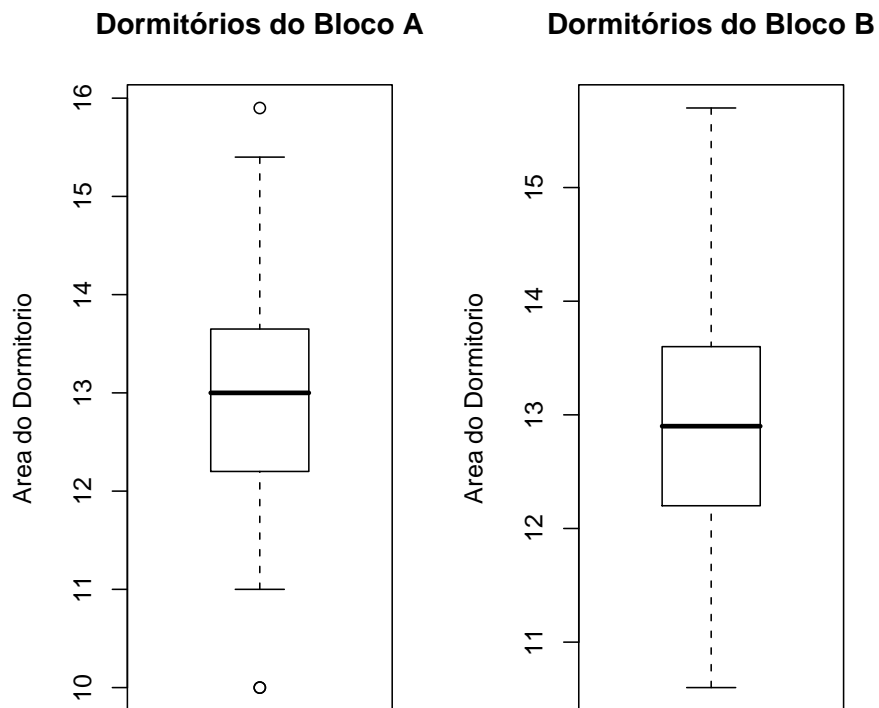

Banheiros do Bloco A



Banheiros do Bloco B



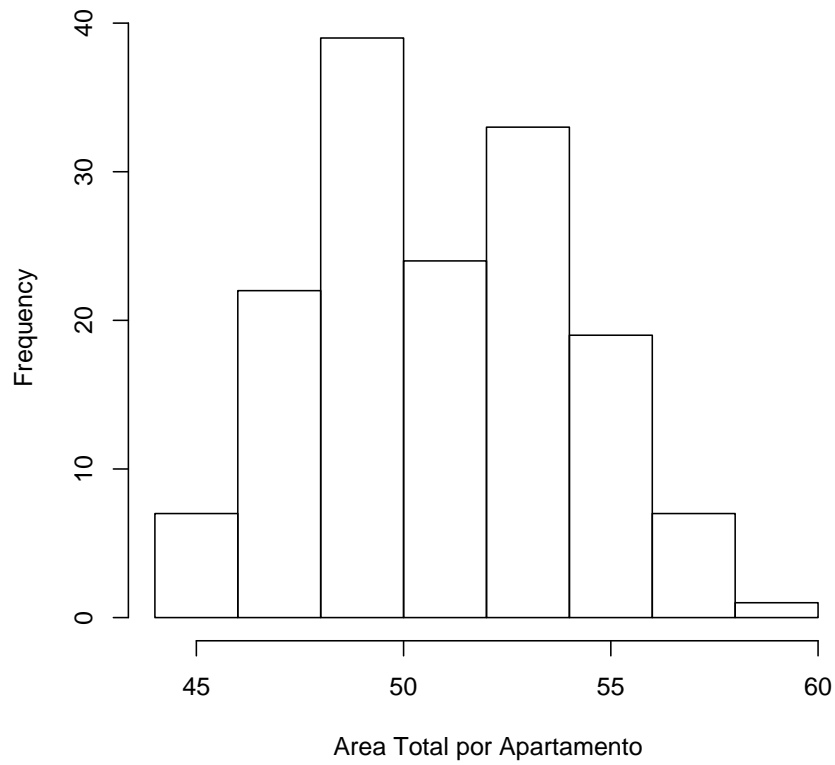
```
> par(mfrow = c(1, 2))  
> boxplot(Dorm[Bloco == "A"], main = "Dormitórios do Bloco A",  
+         ylab = "Area do Dormitorio")  
> boxplot(Dorm[Bloco == "B"], main = "Dormitórios do Bloco B",  
+         ylab = "Area do Dormitorio")
```



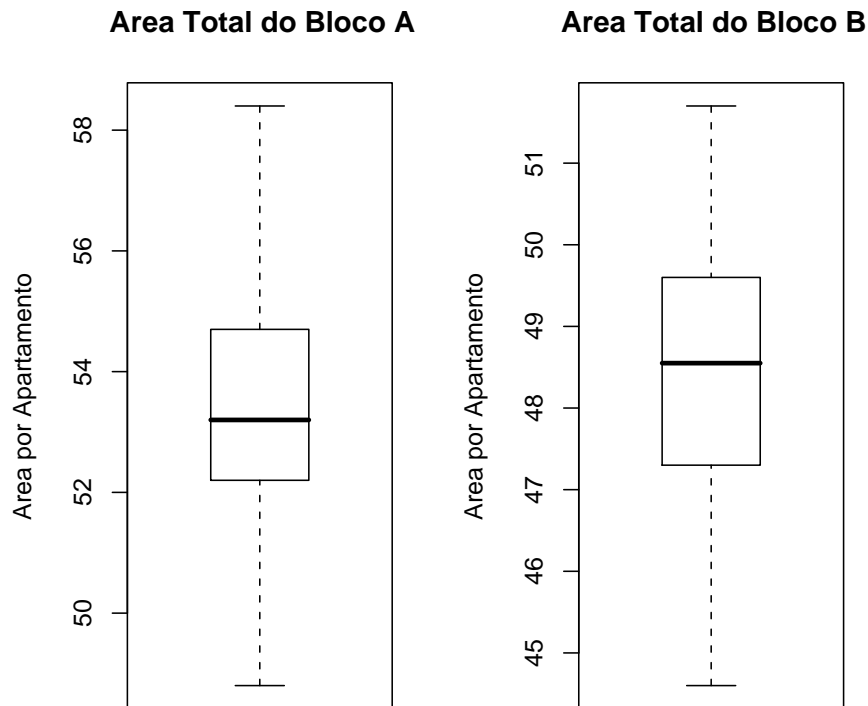
- (d) Calcule a área total para cada apartamento. Armazene esta informação em uma variável denominada Total. Repita os itens (b) e (c) para a variável Total.

```
> Total <- Sala + Cozinha + Banheiro + Dorm
> hist(Total, main = "Histograma da Área Útil", xlab = "Area Total por Apartamento")
```

Histograma da Área Útil



```
> par(mfrow = c(1, 2))  
> boxplot(Total[Bloco == "A"], main = "Area Total do Bloco A",  
+         ylab = "Area por Apartamento")  
> boxplot(Total[Bloco == "B"], main = "Area Total do Bloco B",  
+         ylab = "Area por Apartamento")
```



(e) Baseando-se nos itens anteriores, você diria que existem diferenças nas áreas dos apartamentos dos Blocos A e B? Em caso de positivo, qual(is) cômodo(s) apresenta(m) o problema?

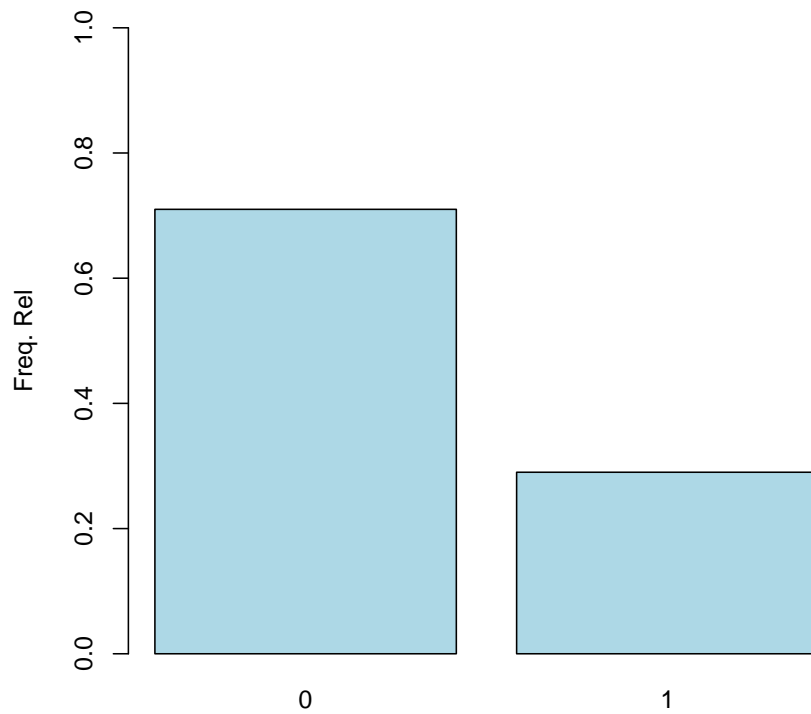
Sim, os cômodos que apresentaram uma variação maior foram a sala e cozinha.

(f) Explore descritivamente os dados referentes a problemas estruturais (rachaduras e infiltrações). Com a informação contida na variável Andar divida os apartamentos em três categorias dependendo do andar onde se encontra: baixo, médio e alto. Estude a ocorrência de rachaduras e infiltrações para cada categoria.

```
> table(Infiltr)
Infiltr
  0   1
108  44

> tab.Infiltr <- round(prop.table(table(Infiltr)), 2)
> tab.Infiltr
Infiltr
  0   1
0.71 0.29

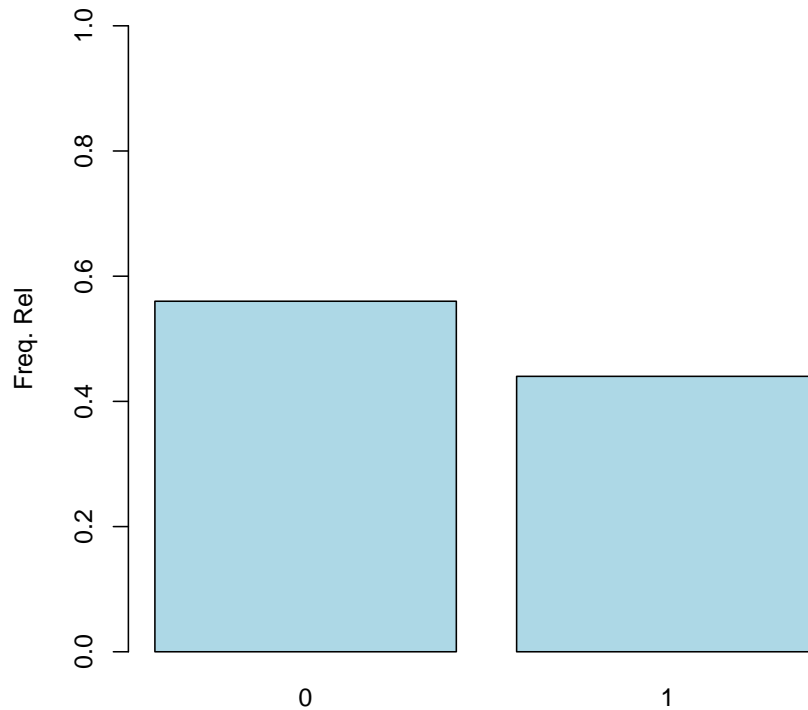
> barplot(tab.Infiltr, col = c("lightblue"), ylim = c(0, 1), ylab = "Freq. Rel")
```



```
> table(Rachadura)
Rachadura
 0  1
85 67

> tab.Rac <- round(prop.table(table(Rachadura)), 2)
> tab.Rac
Rachadura
 0  1
0.56 0.44

> barplot(tab.Rac, col = c("lightblue"), ylim = c(0, 1), ylab = "Freq. Rel")
```



```

> baixo <- Andar <= 6
> medio <- Andar >= 7 & Andar <= 12
> alto <- Andar >= 13
> summary(baixo[Infiltr == 1])
  Mode FALSE  TRUE
logical   37    7
> summary(medio[Infiltr == 1])
  Mode FALSE  TRUE
logical   35    9
> summary(alto[Infiltr == 1])
  Mode FALSE  TRUE
logical   16   28
> summary(baixo[Rachadura == 1])
  Mode FALSE  TRUE
logical   55   12
> summary(medio[Rachadura == 1])
  Mode FALSE  TRUE
logical   31   36
> summary(alto[Rachadura == 1])
  Mode FALSE  TRUE
logical   48   19

```