# UNIVERSIDADE DE SÃO PAULO- USP FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO FÍSICA APLICADA À MEDICINA E BIOLOGIA

# Análise Estatística em Estudos em Saúde Utilizando o Programa R

Lista 2

PROF. EDSON ZANGIACOMI MARTINEZ

Aluna: Stephanie Lisboa Dellevedove

Os seguintes comandos foram utilizados para a importação dos dados enviados pelo professor:

```
> #Importando o banco de dados do computador
> dados<-read.table("C:/Users/Stephanie/Downloads/R/R/antropo.txt",hea
der=T)</pre>
```

#### Exercício 1

O seguinte comando foi utilizado para acrescentar uma coluna aos dados, com o valor do peso convertido para Kg:

```
> #Adicionando uma coluna de dados do peso em Kg(o nome da variavel se
rá weightkg)
> dados$weightkg <- dados$weight*0.45359237</pre>
```

A nova coluna pode ser vista na tela ao pedir para o R mostrar os dados, como pode ser visto na figura abaixo, das primeiras linhas dos dados:

```
> dados
    Density Fat Age weight
1.0708 12.3 23 154.25
1.0853 6.1 22 173.25
              Fat Age Weight Height Neck Chest Abdomen
                                 67.75 36.2
72.25 38.5
                                                                 94.5
98.7
                                                                         59.0 37.3
58.7 37.3
                                                                                               32.0
30.5
                                                                                                              17.1
18.2
                                                93.1
                                                           85.2
                                                                                      21.9
                                                                                                         27.4
                                                                                                                       69.96662
                                                 93.6
                                  66.25 34.0 95.8
72.25 37.4 101.8
     1.0414 25.3
                     22 154.00
                                                           87.9
                                                                  99.2
                                                                         59.6 38.9
                                                                                      24.0
                                                                                               28.8
                                                                                                         25.2
                                                                                                               16.6
                                                                                                                       69.85322
                     26 184.75
                                                           86.4 101.2
                                                                                               32.4
                                                                                                                       83.80119
     1.0751 10.4
                                                                        60.1 37.3
                                                                                      22.8
                                                                                                         29.4
                                                                                                               18.2
                     24 184.25
                                                         100.0 101.9
     1.0502 20.9 24 210.25 74.75 39.0 104.5
                                                           94.4 107.8 66.0 42.0
                                                                                                         30.6 18.8 95.36780
```

**Tabela 1:** Primeiras linhas da tabela de dados, contendo a nova coluna com a variável peso, convertido para kg.

### Exercício 2

O seguinte comando foi utilizado para acrescentar uma coluna aos dados, com o valor da altura convertida para cm:

```
> #Adicionando uma coluna de dados da altura em cm(o nome da variavel
será heightcm)
> dados$heightcm <- dados$Height*2.54</pre>
```

A nova coluna pode ser vista na tela ao pedir para o R mostrar os dados, como pode ser visto na figura abaixo, das primeiras linhas dos dados:

```
> dados
      Density Fat Age Weight Height Neck Chest Abdomen
1.0708 12.3 23 154.25 67.75 36.2 93.1 85.2
1.0853 6.1 22 173.25 72.25 38.5 93.6 83.0
                                                                                                Hip Thigh Knee Ankle Biceps Forearm Wrist
94.5 59.0 37.3 21.9 32.0 27.4 17.1
98.7 58.7 37.3 23.4 30.5 28.9 18.2
                                                                                                                                                                                 weightkg heightcm
69.96662 172.085
        1.0853 6.1 22 173.25
1.0414 25.3 22 154.00
                                                                                                                                                                                 78.58488
                                                                                                                                                                                                   183.515
                                                   66.25 34.0 95.8
                                                                                                             59.6 38.9
                                                                                                  99.2
                                                                                                                                             28.8
                                                                                                                                                                      16.6
                                                                                                                                                                                 69.85322
        1.0751 10.4 26 184.75 72.25 37.4 101.8
1.0340 28.7 24 184.25 71.25 34.4 97.3
1.0502 20.9 24 210.25 74.75 39.0 104.5
                                                                                       86.4 101.2
                                                                                                            60.1 37.3
63.2 42.2
                                                                                                                                22.8
                                                                                                                                             32.4
                                                                                                                                                           29.4
27.7
                                                                                                                                                                     18.2
17.7
                                                                                                                                                                                 83.80119
                                                                                                                                                                                                   183.515
                                                                                     100.0 101.9
                                                                                       94.4 107.8 66.0 42.0 25.6
                                                                                                                                             35.7
                                                                                                                                                                                 95.36780
```

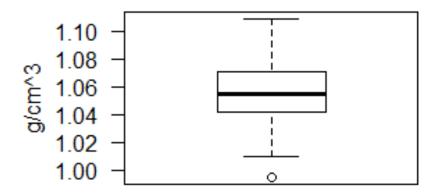
**Tabela 2:** Primeiras linhas da tabela de dados, contendo a nova coluna com a variável altura, convertido para cm.

### Exercício 3

Fazendo um box-plot para cada uma das variáveis contínuas do banco de dados:

> boxplot(dados\$Density,main="Densidade",ylab="g/cm^3",las=2)

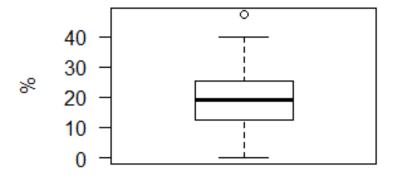
## Densidade



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de um valor atípico, que representa uma pessoa com densidade muito discrepante em relação aos demais ou que houve um erro de medição.

> boxplot(dados\$Fat,main="Percentual de gordura",ylab="%",las=2)

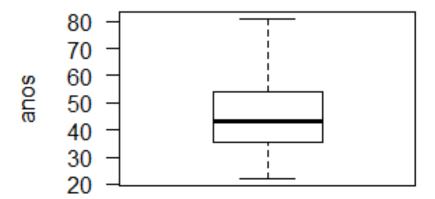
## Percentual de gordura



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de um valor atípico, que representa uma pessoa com densidade de gordura muito discrepante em relação aos demais ou que houve um erro de medição.

> boxplot(dados\$Age,main="Idade",ylab="anos",las=2)

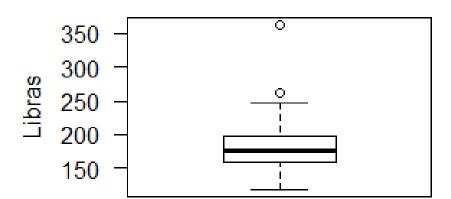
## Idade



Como a mediana (linha mais escura) se encontra deslocada em relação ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma assimétrica. Não há presença de um valor atípico.

> boxplot(dados\$weight,main="Peso",ylab="Libras",las=2)

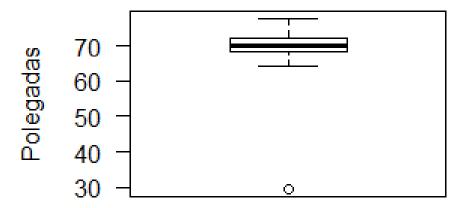




Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com peso muito discrepante em relação aos demais ou que houve um erro de medição.

> boxplot(dados\$Height,main="Altura",ylab="Polegadas",las=2)

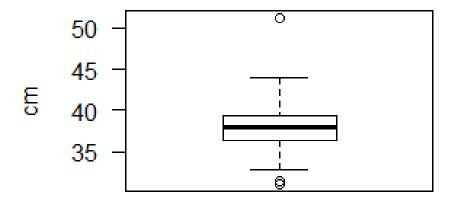
## Altura



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de um valor atípico, que representa uma pessoa com altura muito discrepante em relação aos demais ou que houve um erro de medição.

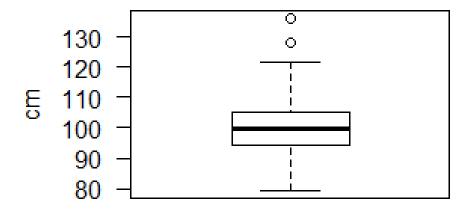
> boxplot(dados\$Neck,main="Circunferência do Pescoço",ylab="cm",las=2)

# Circunferência do Pescoço



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do pescoço muito discrepante em relação aos demais ou que houve um erro de medição.

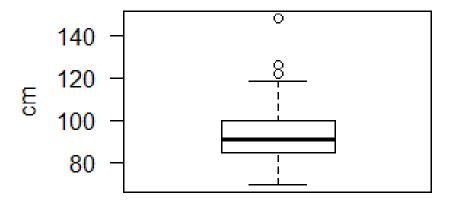
## Circunferência do Peito



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do peito muito discrepante em relação aos demais ou que houve um erro de medição.

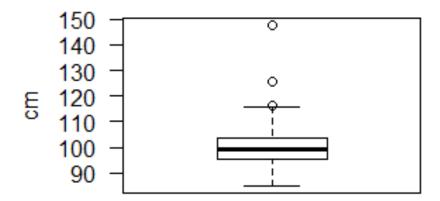
> boxplot(dados\$Abdomen,main="Circunferência do Abdomen",ylab="cm",las =2)

## Circunferência do Abdomen



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do abdômen muito discrepante em relação aos demais ou que houve um erro de medição.

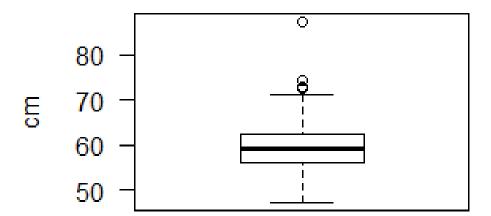
## Circunferência do Quadril



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do quadril muito discrepante em relação aos demais ou que houve um erro de medição.

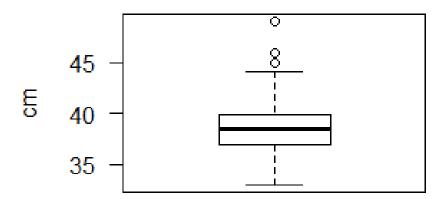
> boxplot(dados\$Thigh,main="Circunferência da Coxa",ylab="cm",las=2)

## Circunferência da Coxa



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência da coxa muito discrepante em relação aos demais ou que houve um erro de medição.

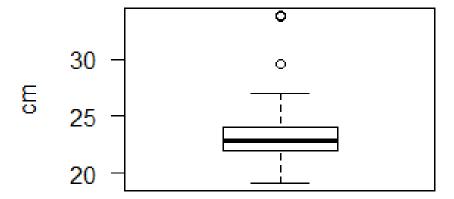
## Circunferência do Joelho



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do joelho muito discrepante em relação aos demais ou que houve um erro de medição.

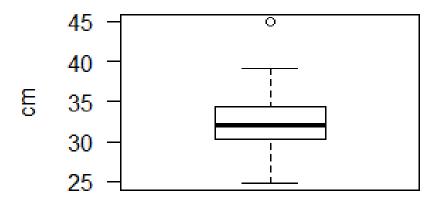
> boxplot(dados\$Ankle,main="Circunferência do Tornozelo",ylab="cm",las =2)

## Circunferência do Tornozelo



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do tornozelo muito discrepante em relação aos demais ou que houve um erro de medição.

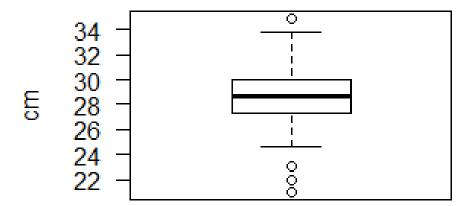
# Circunferência do Bíceps



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de um valor atípico, que representa uma pessoa com circunferência do bíceps muito discrepante em relação aos demais ou que houve um erro de medição.

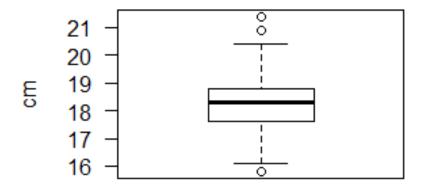
> boxplot(dados\$Forearm,main="Circunferência do Antebraço",ylab="cm",l
as=2)

# Circunferência do Antebraço



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do antebraço muito discrepante em relação aos demais ou que houve um erro de medição.

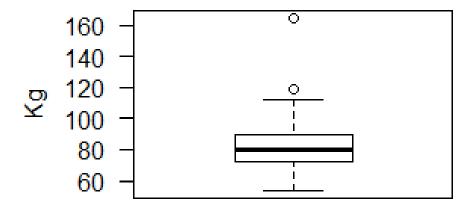
## Circunferência do Pulso



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com circunferência do pulso muito discrepante em relação aos demais ou que houve um erro de medição.

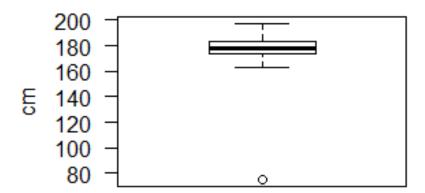
> boxplot(dados\$weightkg,main="Peso (kg)",ylab="Kg",las=2)

# Peso (kg)



Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de valores atípicos, que representam pessoas com peso muito discrepante em relação aos demais ou que houve um erro de medição.

# Altura (cm)

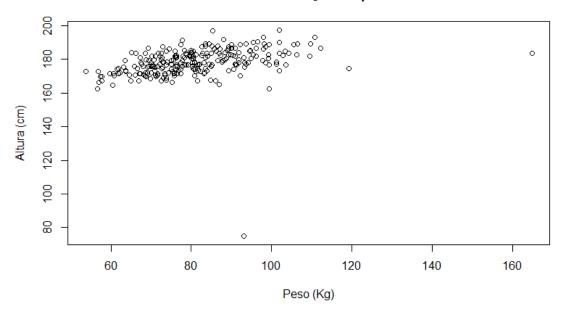


Como a mediana (linha mais escura) se encontra próxima ao centro do conjunto de dados amostrais, há indícios de que os dados estão distribuídos de forma simétrica. Há a presença de um valor atípico, que representa uma pessoa com altura muito discrepante em relação aos demais ou que houve um erro de medição.

### Exercício 4

- > #Construindo o gráfico de dispersão entre o peso e a altura das pess oas
- > plot(dados\$weightkg,dados\$heightcm,ylab="Altura (cm)",xlab="Peso (Kg
  )",main = "Altura em função do peso")

### Altura em função do peso



No gráfico de dispersão da altura em função do peso é possível ver que a maioria dos dados se encontram concentrados em uma mesma região, tendo um comportamento

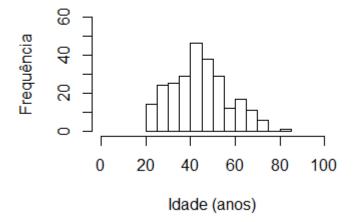
levemente crescente; como era de se esperar uma relação entre altura e peso para pessoas; entretanto há alguns valores atípicos:

- Uma pessoa com altura de 74,93cm e peso de 92,98kg. Considerando o conhecimento de medidas consideradas dentro da normalidade para pessoas, é provável que a medida tenha sido anotada de forma equivocada (a pessoa que fez as medidas pode se certificar se esse foi um erro de medida ou se os dados para essa pessoa realmente são verídicos).
- Uma pessoa com altura de 183,52cm e peso de 164,72kg. Possuindo assim uma massa muito maior que as demais pessoas (a pessoa que fez as medidas pode se certificar se esse foi um erro de medida ou se os dados para essa pessoa realmente são verídicos).

### Exercício 5

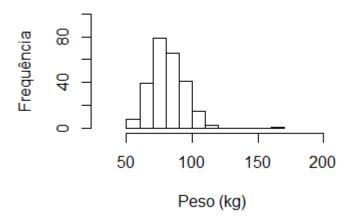
```
> hist(dados$Age, main="Histograma das Idades", xlab="Idade (anos)",ylab="Frequência", ylim = c(0,60),xlim = c(0,100))
```

### Histograma das Idades



A partir do histograma é possível notar que os dados foram colhidos de uma população adulta, e que nessa faixa de idade, a partir dos 20 anos, os dados se mostram aproximadamente simétricos em torno da faixa de 40 anos, mostrando que há uma concentração maior de pessoas na faixa etária em torno de 40 anos

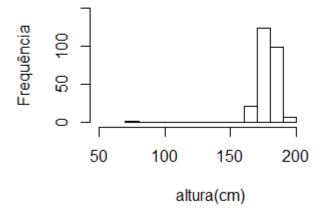
## Histograma dos Pesos



Considerando que existe um valor atípico nesse conjunto de dados, ao analisar o conjunto de dados excluindo esse valor atípico é possível perceber que os dados estão distribuídos de uma forma aproximadamente simétrica em torno de um valor próximo a 70kg.

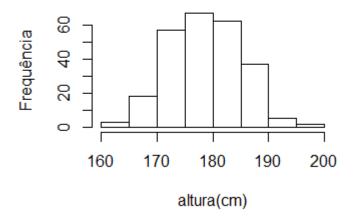
> hist(dados\$heightcm,main="Histograma das alturas",xlab="altura(cm)", ylab="Frequência", ylim = c(0,150),xlim = c(50,220))

### Histograma das alturas



Observando o histograma é possível ver que há um valor que se destaca muito da distribuição dos dados. Desconsiderando esse valor, que pode ter sido um erro de anotação, o histograma fica como na imagem a seguir:

## Histograma das alturas



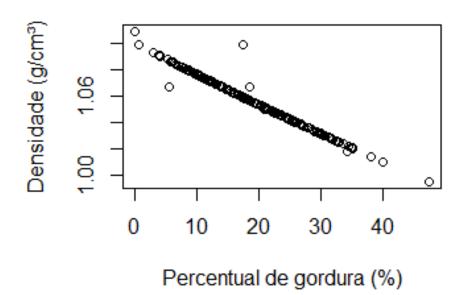
Analisando esse histograma é possível ver que as alturas das pessoas da pesquisa estão distribuídas de forma aproximadamente simétrica em torno de um valor em torno de 180cm, o que é muito próximo à média, considerando esses dados apenas (178,58cm)

### Exercício 6

Calculando os gráficos de dispersão da gordura m função das outras variáveis antropométricas

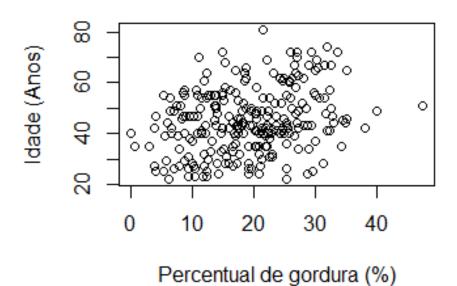
> plot(dados\$Fat,dados\$Density,xlab="Percentual de gordura (%)",ylab=" Densidade (g/cm $^3$ )",main = "Densidade")

## Densidade

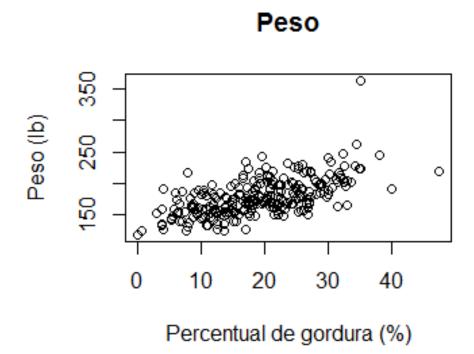


> plot(dados\$Fat,dados\$Age,xlab="Percentual de gordura (%)",ylab="Idad
e (Anos)",main = "Idade")

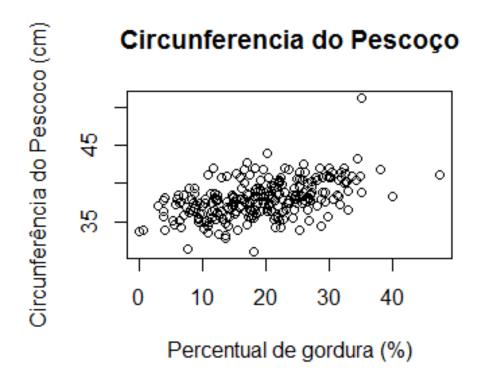
## Idade



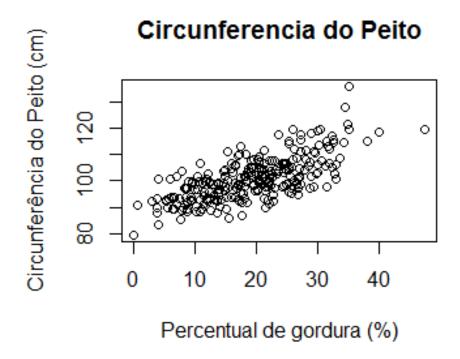
> plot(dados\$Fat,dados\$Weight,xlab="Percentual de gordura (%)",ylab="P
eso (lb)",main = "Peso")



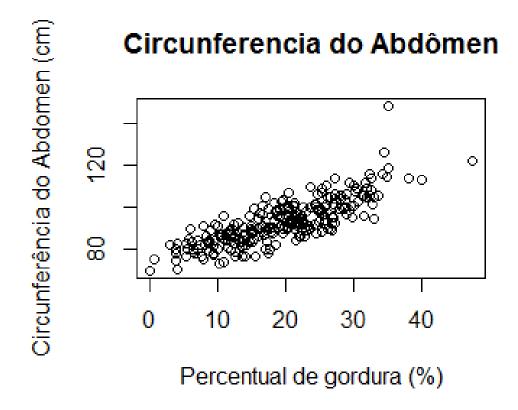
> plot(dados\$Fat,dados\$Neck,xlab="Percentual de gordura (%)",ylab="Cir cunferência do Pescoco (cm)",main = "Circunferencia do Pescoço")



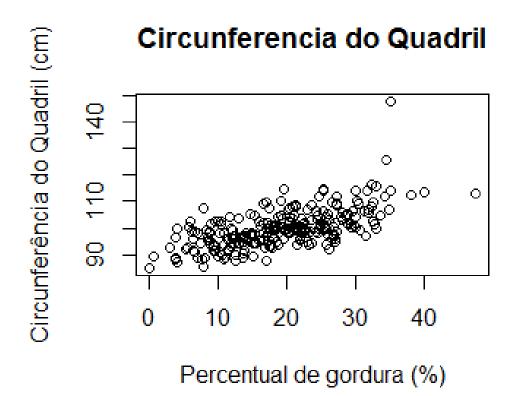
> plot(dados\$Fat,dados\$Chest,xlab="Percentual de gordura (%)",ylab="Ci rcunferência do Peito (cm)",main = "Circunferencia do Peito")

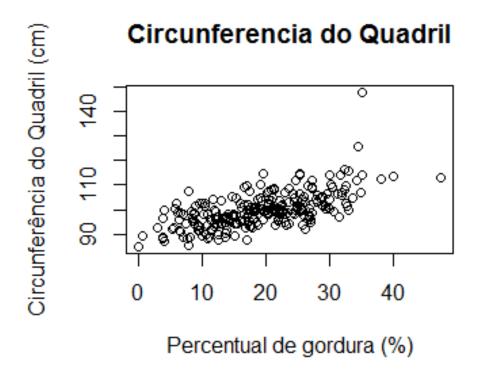


> plot(dados\$Fat,dados\$Abdomen,xlab="Percentual de gordura (%)",ylab="
Circunferência do Abdomen (cm)",main = "Circunferencia do Abdômen")

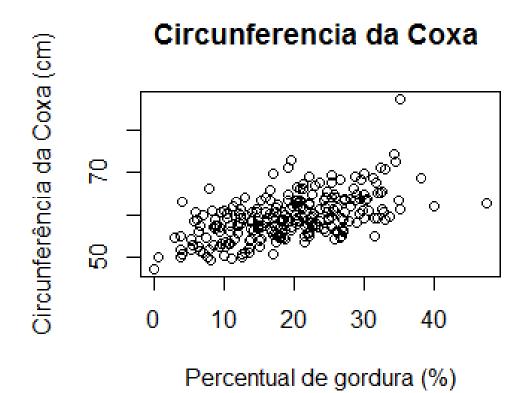


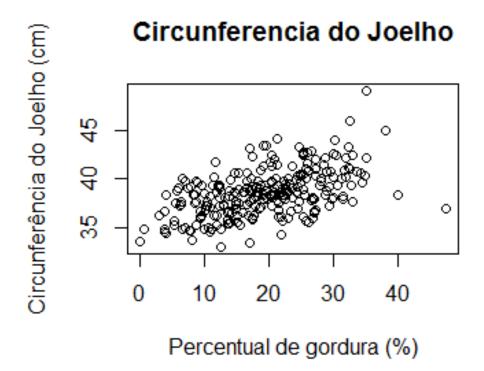
> plot(dados\$Fat,dados\$Hip,xlab="Percentual de gordura (%)",ylab="Circ
unferência do Quadril (cm)",main = "Circunferencia do Quadril")



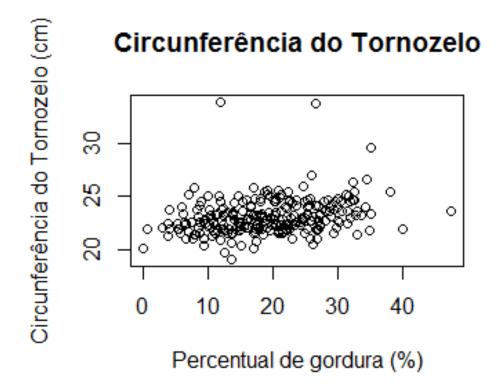


> plot(dados\$Fat,dados\$Thigh,xlab="Percentual de gordura (%)",ylab="Ci
rcunferência da Coxa (cm)",main = "Circunferencia da Coxa ")

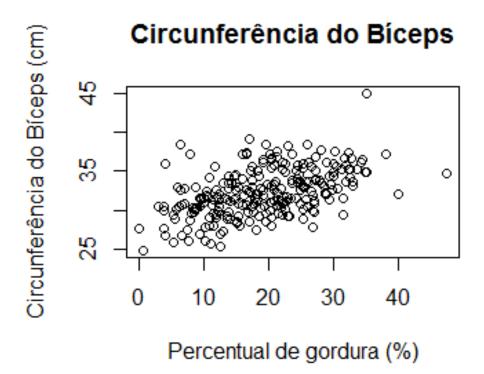




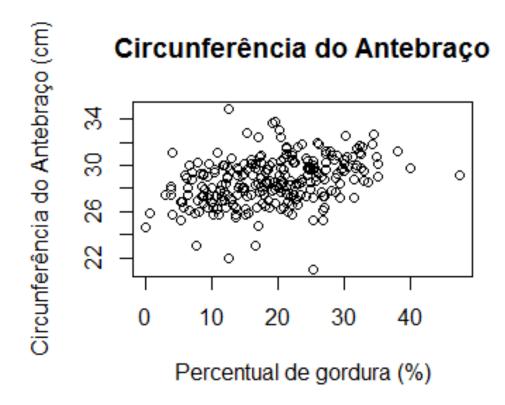
> plot(dados\$Fat,dados\$Ankle,xlab="Percentual de gordura (%)",ylab="Ci
rcunferência do Tornozelo (cm)",main = "Circunferência do Tornozelo")



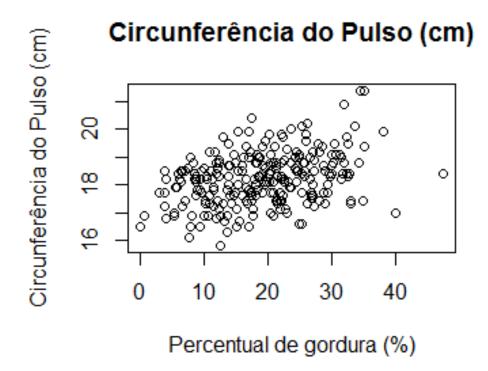
> plot(dados\$Fat,dados\$Biceps,xlab="Percentual de gordura (%)",ylab="C
ircunferência do Bíceps (cm)",main = "Circunferência do Bíceps")



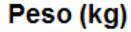
> plot(dados\$Fat,dados\$Forearm,xlab="Percentual de gordura (%)",ylab="
Circunferência do Antebraço (cm)",main = "Circunferência do Antebraço"
)

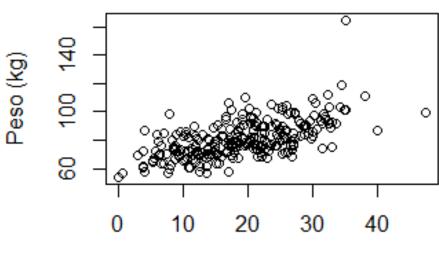


> plot(dados\$Fat,dados\$Wrist,xlab="Percentual de gordura (%)",ylab="Ci
rcunferência do Pulso (cm)",main = "Circunferência do Pulso (cm)")



> plot(dados\$Fat,dados\$weightkg,xlab="Percentual de gordura (%)",ylab=
"Peso (kg)",main = "Peso (kg)")

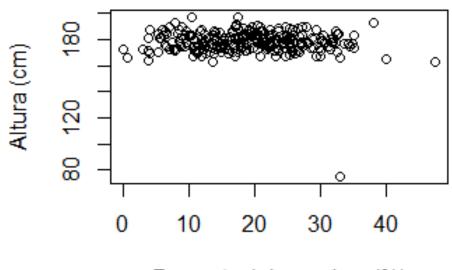




Percentual de gordura (%)

> plot(dados\$Fat,dados\$heightcm,xlab="Percentual de gordura (%)",ylab=
"Altura (cm)",main = "Altura (cm)")

# Altura (cm)



# Percentual de gordura (%)

### Exercício 7

```
> cor.test(dados$Fat,dados$Density,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Density
t = -100.22, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9904570 -0.9843641
sample estimates:
-0.9877824
> cor.test(dados$Fat,dados$Age,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Age
t = 4.8175, df = 250, p-value = 2.522e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1741581 0.4006030
sample estimates:
      cor
0.2914584
> cor.test(dados$Fat,dados$weight,method = c("pearson"))
       Pearson's product-moment correlation
```

```
data: dados$Fat and dados$Weight
t = 12.249, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5288644 0.6842076
sample estimates:
     cor
0.612414
> cor.test(dados$Fat,dados$Height,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Height
t = -1.4207, df = 250, p-value = 0.1566
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.21073764 0.03445855
sample estimates:
        cor
-0.08949538
> cor.test(dados$Fat,dados$Neck,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Neck
t = 8.9018, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3907050 0.5790599
sample estimates:
      cor
0.4905919
> cor.test(dados$Fat,dados$Chest,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Chest
t = 15.613, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.6341034 0.7601899
sample estimates:
      cor
0.7026203
> cor.test(dados$Fat,dados$Abdomen,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Abdomen
t = 22.112, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7669520 0.8514218
sample estimates:
      cor
0.8134323
```

```
> cor.test(dados$Fat,dados$Hip,method = c("pearson"))
        Pearson's product-moment correlation
data: dados$Fat and dados$Hip
t = 12.666, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5436276 0.6950738
sample estimates:
      cor
0.6252009
> cor.test(dados$Fat,dados$Thigh,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Thigh
t = 10.676, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4684275 0.6389926
sample estimates:
      cor
0.5596075
> cor.test(dados$Fat,dados$Knee,method = c("pearson"))
        Pearson's product-moment correlation
data: dados$Fat and dados$Knee
t = 9.3415, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4109218 0.5948476
sample estimates:
      cor
0.5086652
> cor.test(dados$Fat,dados$Ankle,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Ankle
t = 4.3625, df = 250, p-value = 1.882e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1472361 0.3771471
sample estimates:
      cor
0.2659698
> cor.test(dados$Fat,dados$Biceps,method = c("pearson"))
        Pearson's product-moment correlation
data: dados$Fat and dados$Biceps
t = 8.966, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3936960 0.5814045
sample estimates:
```

```
cor
0.4932711
> cor.test(dados$Fat,dados$Forearm,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Forearm
t = 6.1282, df = 250, p-value = 3.437e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2489307 0.4642284
sample estimates:
      cor
0.3613869
> cor.test(dados$Fat,dados$Wrist,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$Wrist
t = 5.8419, df = 250, p-value = 1.6e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2329799 0.4508395
sample estimates:
      cor
0.3465749
> cor.test(dados$Fat,dados$weightkg,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$weightkg
t = 12.249, df = 250, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5288644 0.6842076
sample estimates:
     cor
0.612414
> cor.test(dados$Fat,dados$heightcm,method = c("pearson"))
       Pearson's product-moment correlation
data: dados$Fat and dados$heightcm
t = -1.4207, df = 250, p-value = 0.1566
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.21073764 0.03445855
sample estimates:
        cor
-0.08949538
```

Considerando todas as variáveis, a variável antropométrica que tem uma correlação maior com o percentual de gordura é a densidade, tendo um coeficiente de correlação de 0,9877824, muito próximo de 1, que seria o máximo valor possível, caso houvesse

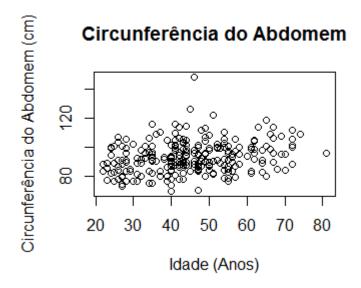
uma correlação perfeita entre os dados, mostrando que há uma correlação muito alta entre os dados.

#### Exercício 8

Para ter uma noção de como os dados se relacionam, foi feito um gráfico da dispersão da circunferência do abdômen em função da idade, utilizando o seguinte código:

```
> plot(dados$Age,dados$Abdomen,xlab="Idade (Anos)",ylab="Circunferênci
a do Abdomem (cm)",main = "Circunferência do Abdomem")
```

O resultado é visto na figura a seguir:



Pelo gráfico é difícil afirmar se há alguma correlação entre as duas variáveis, portanto foi feito um teste de correlação de Pearson com essas variáveis:

```
Pearson's product-moment correlation

data: dados$Abdomen and dados$Age
t = 3.7438, df = 250, p-value = 0.0002249
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
    0.1099676    0.3441826
sample estimates:
        cor
0.2304094
```

> cor.test(dados\$Abdomen,dados\$Age,method = c("pearson"))

Como o resultado para o teste foi de um coeficiente de correlação de 0,2304094; um valor bem pequeno, pode-se concluir não há indícios de que haja uma correlação entre essas a idade e a circunferência do abdomem.