



## Solução de Atividade da Extensão dia 06/12/2019

---

Giovani Nucci, Caroline Assis

---

Departamento de Estatística

**Áreas do Conhecimento:**

Estatística Aplicada, R

**Palavras-Chave:**

regressao linear, estatística descritiva,  
correlação, gráficos

Este paper é a Solução das atividades do dia 06/12/2019 da turma de extensão.

## 1 Primeiros - Passos

Primeiramente iremos verificar se os pacotes que utilizaremos estão baixados, caso negativo os baixamos e depois carregamos para podermos usar as funções.

```
library(readr)
library(fdth)
library(OneTwoSamples)
library(TeachingDemos)
```

## 2 Lista 1

Importamos o banco de Dados chamado de *bd\_ex\_0612.txt* From text(base).

```
caminho <- file.choose()

banco <- read.delim(caminho)

View(banco)

str(banco)

## 'data.frame': 30 obs. of 4 variables:
## $ Empregado : int 1 2 3 4 5 6 7 8 9 10 ...
## $ salarios_anuais : int 41200 39565 30980 23225 21250 41875 31225 30135 29850 21850 ...
## $ idade : int 45 43 41 27 26 45 41 36 32 22 ...
## $ estado_de_emprego: int 2 1 1 1 1 2 1 2 1 1 ...

attach(banco)
```

### 2.1 a) Transforme a variável Estado de Emprego em um fator, em que 1 - Ativo e 2 - Demitido.

```
estado_de_emprego <- factor(x = estado_de_emprego,
                             labels = c("Ativo", "Demitido"),
                             levels = c(1,2))
```

### 2.2 b) Construa uma tabela completa de frequência com a variável estado de emprego.

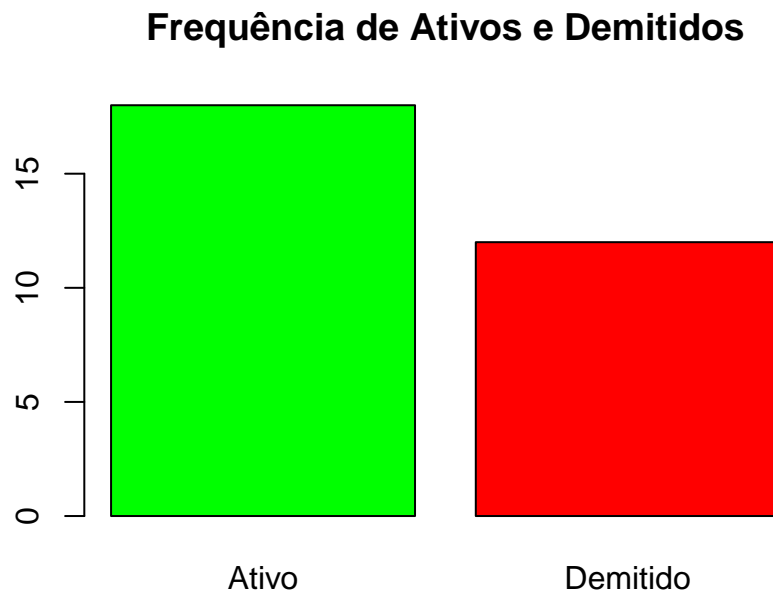
```
tabela_1 <- fdt(as.numeric(estado_de_emprego),
                 start = 0,
                 end = 2,
                 right = TRUE,
                 h = 1)

tabela_1

## Class limits f rf rf(%) cf cf(%)
## (0,1] 18 0.6 60 18 60
## (1,2] 12 0.4 40 30 100
```

2.3 c) Construa um gráfico de barras ou colunas para a variável estado de emprego.

```
barplot(height = table(estado_de_emploi),  
        col = c("green", "red"),  
        main = "Frequência de Ativos e Demitidos")
```



2.4 d) Gere um gráfico de setores para a variável estado de emprego.

```
pie(x = table(estado_de_emprego),  
    col = c("green", "red"),  
    main = "Percentual de Empregados e Desempregados",  
    labels = c("Ativo", "Demitido"))
```

### Percentual de Empregados e Desempregados



2.5 e) Construa uma tabela de frequência por classes com a variável salário.

```
tabela_2 <- fdt(x = salarios_anuais,  
                breaks = "Sturges")  
tabela_2
```

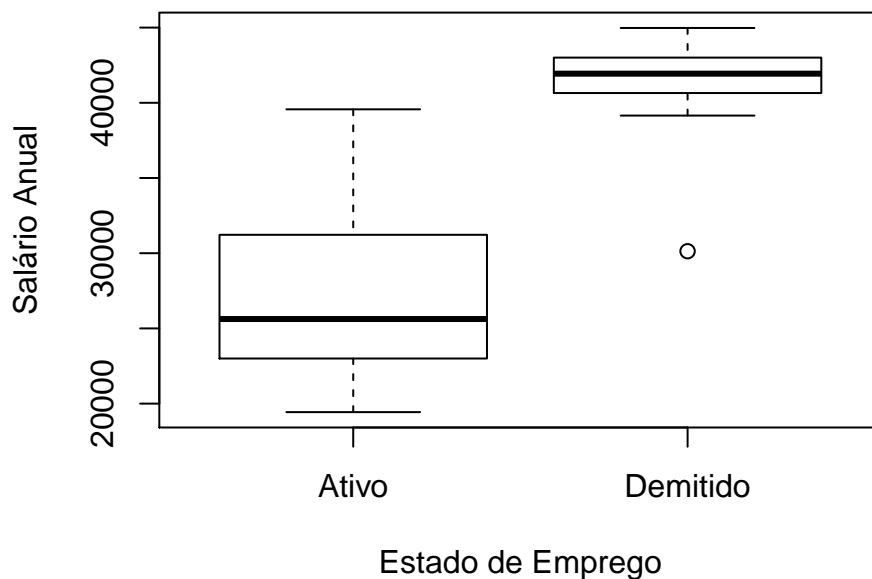
##	Class	limits	f	rf	rf(%)	cf	cf(%)
##	[19241,	23605)	6	0.20	20.00	6	20.00
##	[23605,	27969)	4	0.13	13.33	10	33.33
##	[27969,	32333)	6	0.20	20.00	16	53.33
##	[32333,	36697)	1	0.03	3.33	17	56.67
##	[36697,	41061)	5	0.17	16.67	22	73.33
##	[41061,	45425)	8	0.27	26.67	30	100.00

2.6 f) Construa uma tabela de frequência por classes com a variável idade.

```
tabela_3 <- fdt(x = idade,  
               breaks = "Sturges")  
tabela_3  
  
##   Class limits f    rf rf(%) cf   cf(%)  
##      [19,25) 4 0.13 13.33  4  13.33  
##      [25,31) 6 0.20 20.00 10  33.33  
##      [31,37) 2 0.07  6.67 12  40.00  
##      [37,43) 8 0.27 26.67 20  66.67  
##      [43,49) 8 0.27 26.67 28  93.33  
##      [49,56) 2 0.07  6.67 30 100.00
```

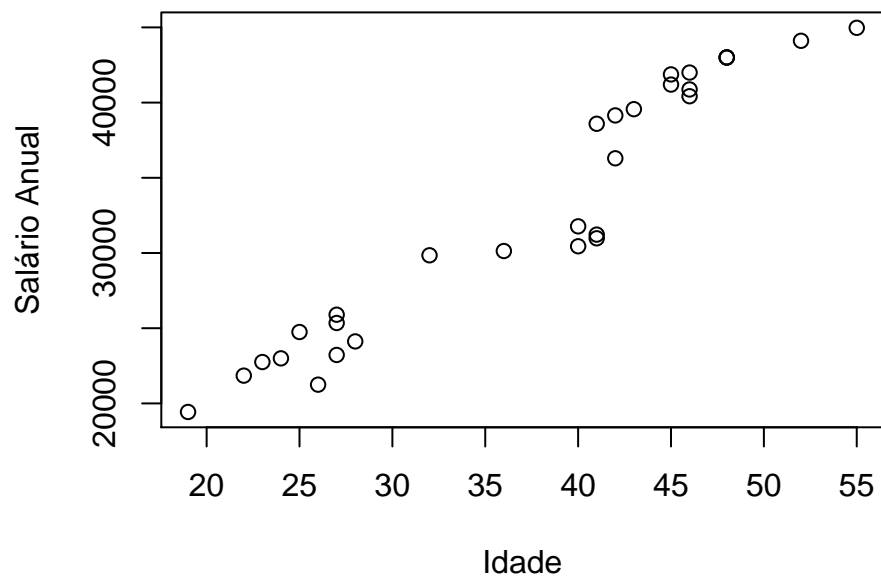
2.7 g) Gere um gráfico boxplot para a variável salário por estado de emprego.

```
boxplot(salarios_anuais ~ estado_de_emprego,  
        xlab = "Estado de Emprego",  
        ylab = "Salário Anual")
```



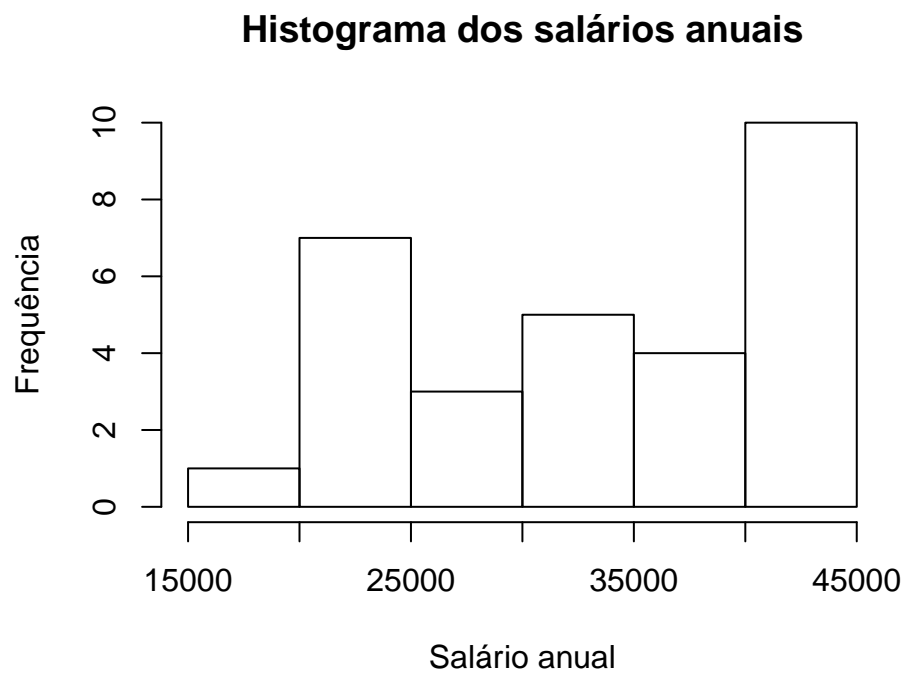
2.8 h) Gere um gráfico de dispersão para a variável salário por idade.

```
plot(salarios_anuais ~ idade,  
     xlab = "Idade",  
     ylab = "Salário Anual")
```



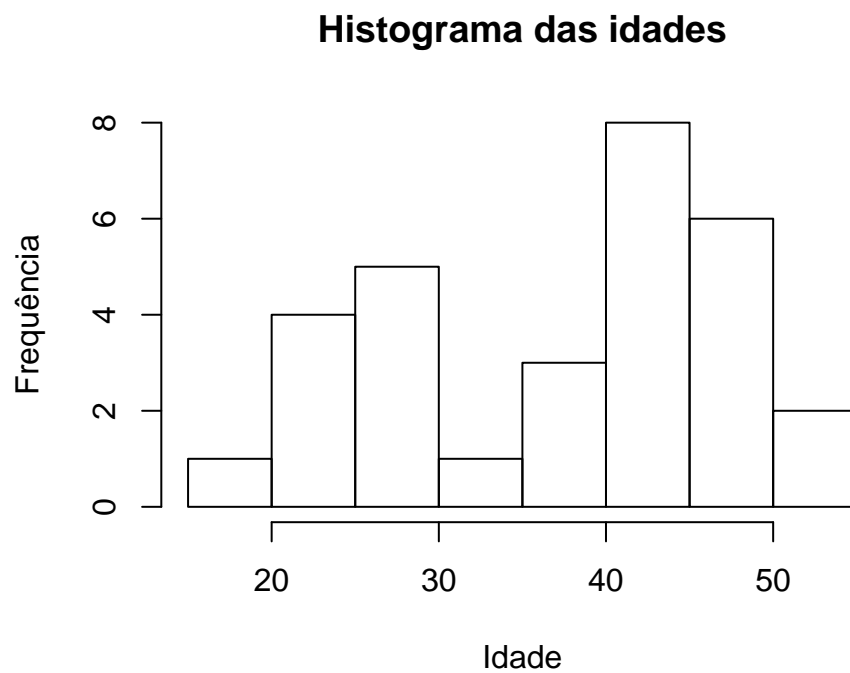
2.9 i) Gere um histograma para a variável salário.

```
hist(salarios_anuais,  
     main = "Histograma dos salários anuais",  
     xlab = "Salário anual",  
     ylab = "Frequência")
```



2.10 j) Gere um histograma para a variável idade.

```
hist(idade,  
     main = "Histograma das idades",  
     xlab = "Idade",  
     ylab = "Frequência")
```



2.11 k) Calcule as medidas de tendência central e de dispersão da variável salário.

```
summary(salarios_anuais)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##  19435   24900   31500   33138   41119   44975
```

```
sd(salarios_anuais)
```

```
## [1] 8447.938
```



2.12 l) Calcule as medidas de tendência central e de dispersão da variável idade.

```
summary(idade)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   19.00   27.00   41.00   37.50   45.75   55.00 

sd(idade)

## [1] 10.19381
```

2.13 m) Calcule a idade média dos funcionários com relação ao estado de emprego

```
tapply(idade, estado_de_emprego, FUN = mean)

##      Ativo Demitido 
## 31.55556 46.41667 

#ou

mean(idade[estado_de_emprego == "Ativo"])

## [1] 31.55556

mean(idade[estado_de_emprego == "Demitido"])

## [1] 46.41667
```

2.14 n) Mensure o grau de correlação entre as variáveis salário e idade. Qual o tipo de correlação entre as variáveis?

```
cor(salarios_anuais, idade)

## [1] 0.9629456
```

2.15 o) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(salarios_anuais,idade)

##
## Pearson's product-moment correlation
##
## data: salarios_anuais and idade
## t = 18.893, df = 28, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9228212 0.9824007
## sample estimates:
## cor
## 0.9629456
```

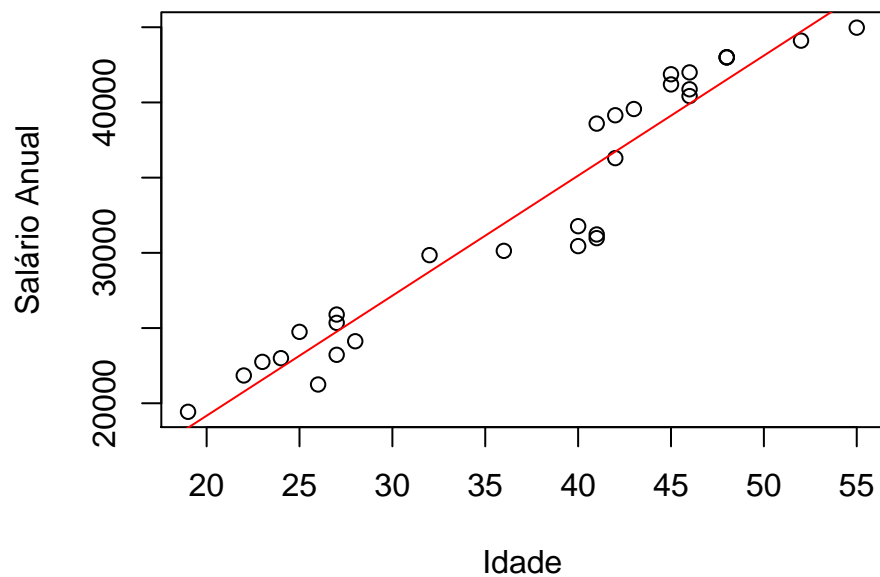
2.16 p) Calcule os coeficientes de regressão.

```
coef <- lm(salarios_anuais~idade)
coef

##
## Call:
## lm(formula = salarios_anuais ~ idade)
##
## Coefficients:
## (Intercept)      idade
##      3212         798
```

### 2.17 q) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(salarios_anuais ~ idade,  
     xlab = "Idade",  
     ylab = "Salário Anual")  
abline(coef, col = "red")
```



### 2.18 r) Existe correlação?

```
cor.test(salarios_anuais,idade)  
  
##  
## Pearson's product-moment correlation  
##  
## data: salarios_anuais and idade  
## t = 18.893, df = 28, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9228212 0.9824007  
## sample estimates:  
## cor  
## 0.9629456
```

## 2.19 s) O modelo está bem ajustado?

```
summary(coef)

##
## Call:
## lm(formula = salarios_anuais ~ idade)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4951  -1508   1008   1488   2752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3211.77    1639.54   1.959   0.0601 .
## idade        798.02     42.24  18.893  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2319 on 28 degrees of freedom
## Multiple R-squared:  0.9273, Adjusted R-squared:  0.9247
## F-statistic: 357 on 1 and 28 DF, p-value: < 2.2e-16
```

### 3 Lista 2

Importando o banco como csv.

```
caminho_2 <- file.choose()

banco_2 <- read_csv(caminho_2)

View(banco_2)

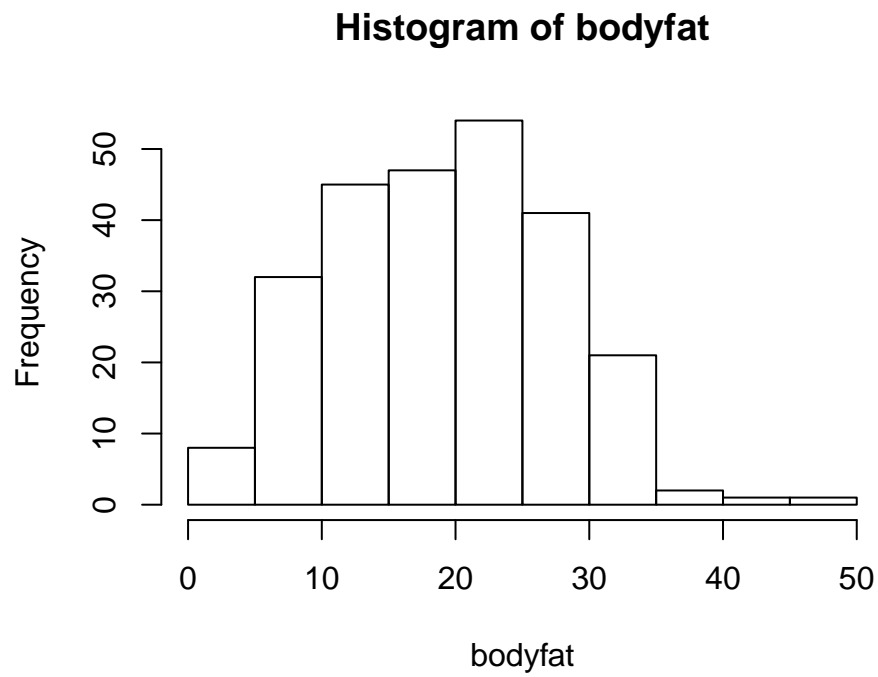
str(banco_2)

## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 252 obs. of  15 variables:
## $ Density: num  1.07 1.09 1.04 1.08 1.03 ...
## $ bodyfat: num  12.3 6.1 25.3 10.4 28.7 20.9 19.2 12.4 4.1 11.7 ...
## $ Age    : num  23 22 22 26 24 24 26 25 25 23 ...
## $ Weight : num  154 173 154 185 184 ...
## $ Height : num  67.8 72.2 66.2 72.2 71.2 ...
## $ Neck   : num  36.2 38.5 34 37.4 34.4 39 36.4 37.8 38.1 42.1 ...
## $ Chest  : num  93.1 93.6 95.8 101.8 97.3 ...
## $ Abdomen: num  85.2 83 87.9 86.4 100 94.4 90.7 88.5 82.5 88.6 ...
## $ Hip    : num  94.5 98.7 99.2 101.2 101.9 ...
## $ Thigh  : num  59 58.7 59.6 60.1 63.2 66 58.4 60 62.9 63.1 ...
## $ Knee   : num  37.3 37.3 38.9 37.3 42.2 42 38.3 39.4 38.3 41.7 ...
## $ Ankle  : num  21.9 23.4 24 22.8 24 25.6 22.9 23.2 23.8 25 ...
## $ Biceps : num  32 30.5 28.8 32.4 32.2 35.7 31.9 30.5 35.9 35.6 ...
## $ Forearm: num  27.4 28.9 25.2 29.4 27.7 30.6 27.8 29 31.1 30 ...
## $ Wrist  : num  17.1 18.2 16.6 18.2 17.7 18.8 17.7 18.8 18.2 19.2 ...
## - attr(*, "spec")=
## .. cols(
## ..   Density = col_double(),
## ..   bodyfat = col_double(),
## ..   Age = col_double(),
## ..   Weight = col_double(),
## ..   Height = col_double(),
## ..   Neck = col_double(),
## ..   Chest = col_double(),
## ..   Abdomen = col_double(),
## ..   Hip = col_double(),
## ..   Thigh = col_double(),
## ..   Knee = col_double(),
## ..   Ankle = col_double(),
## ..   Biceps = col_double(),
## ..   Forearm = col_double(),
## ..   Wrist = col_double()
## .. )

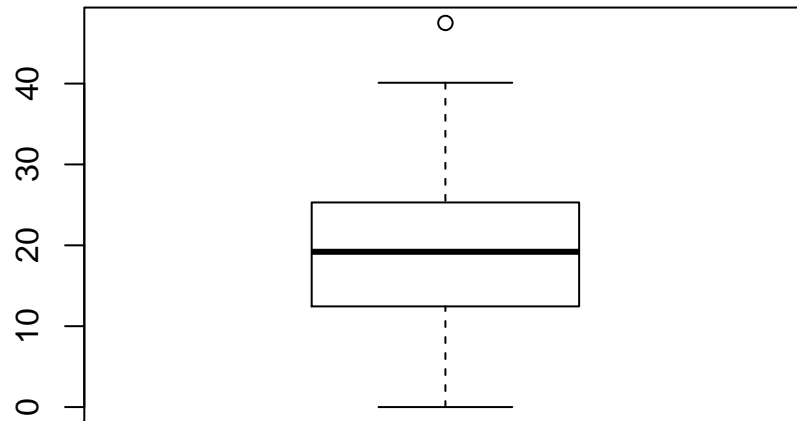
attach(banco_2)
```

3.0.1 1) Faça o histograma e o boxplot da variável Bodyfat.

```
hist(bodyfat)
```

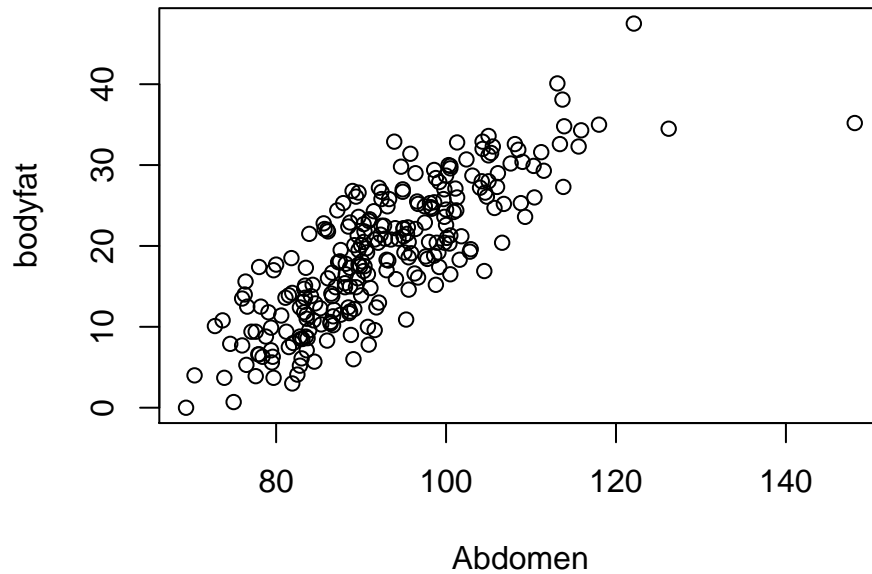


```
boxplot (bodyfat)
```



3.0.2 2) Faça um gráfico de dispersão entre Bodyfat versus Abdomen.

```
plot(bodyfat ~ Abdomen)
```



3.0.3 3) Existe correlação? Qual o tipo de correlação entre as variáveis?

```
cor(bodyfat, Abdomen)
```

```
## [1] 0.8134323
```

3.0.4 4) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(bodyfat, Abdomen)
```

```
##
## Pearson's product-moment correlation
##
## data: bodyfat and 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
```



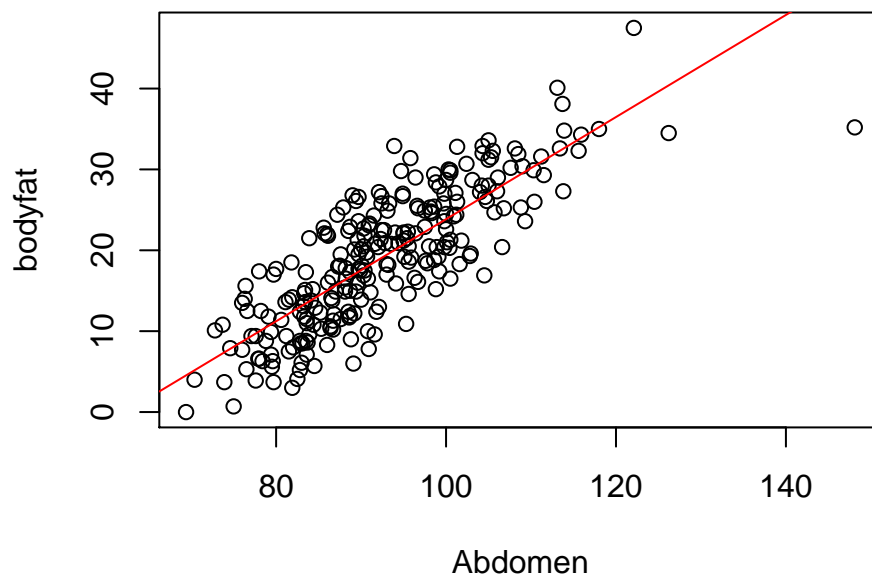
### 3.0.5 5) Calcule os coeficientes de regressão.

```
coef_2 <- lm(bodyfat ~ Abdomen)
coef_2

##
## Call:
## lm(formula = bodyfat ~ Abdomen)
##
## Coefficients:
## (Intercept)      Abdomen
##    -39.2802      0.6313
```

### 3.0.6 6) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(bodyfat ~ Abdomen)
abline(coef_2,col = "red")
```



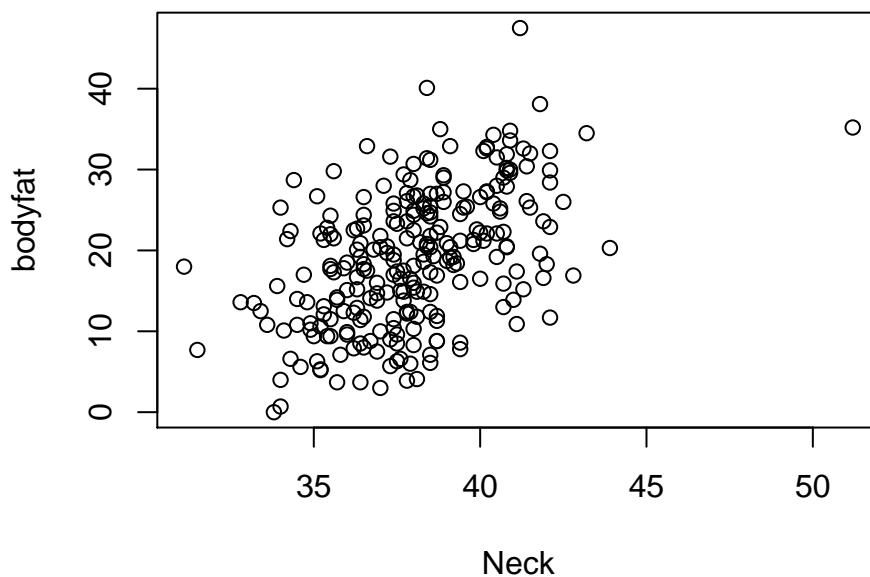
### 3.0.7 7) O modelo está bem ajustado?

```
summary(coef_2)
```

```
##
## Call:
## lm(formula = bodyfat ~ Abdomen)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.0160  -3.7557   0.0554   3.4215  12.9007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -39.28018     2.66034  -14.77  <2e-16 ***
## Abdomen       0.63130     0.02855   22.11  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.877 on 250 degrees of freedom
## Multiple R-squared:  0.6617, Adjusted R-squared:  0.6603
## F-statistic: 488.9 on 1 and 250 DF,  p-value: < 2.2e-16
```

### 3.0.8 8) Faça um gráfico de dispersão entre Bodyfat versus Neck.

```
plot(bodyfat~Neck)
```



### 3.0.9 9) Existe correlação? Qual o tipo de correlação entre as variáveis?

```
cor(bodyfat, Neck)
```

```
## [1] 0.4905919
```

### 3.0.10 10) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(bodyfat, Neck)
```

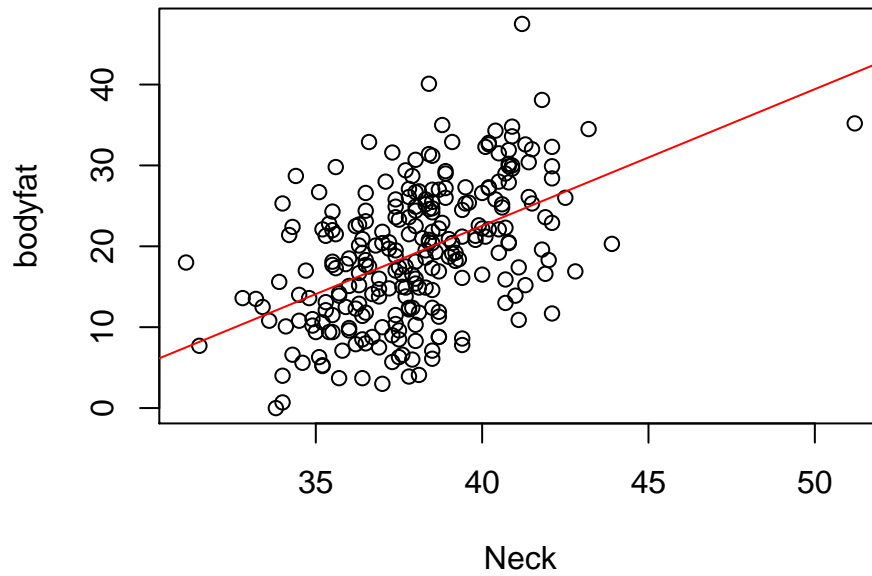
```
##  
## Pearson's product-moment correlation  
##  
## data: bodyfat and Neck  
## t = 8.9018, df = 250, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3907050 0.5790599  
## sample estimates:  
## cor  
## 0.4905919
```

### 3.0.11 11) Calcule os coeficientes de regressão.

```
coef_3 <- lm(bodyfat ~ Neck)
```

3.0.12 12) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(bodyfat~Neck)  
abline(coef_3,col = "red")
```



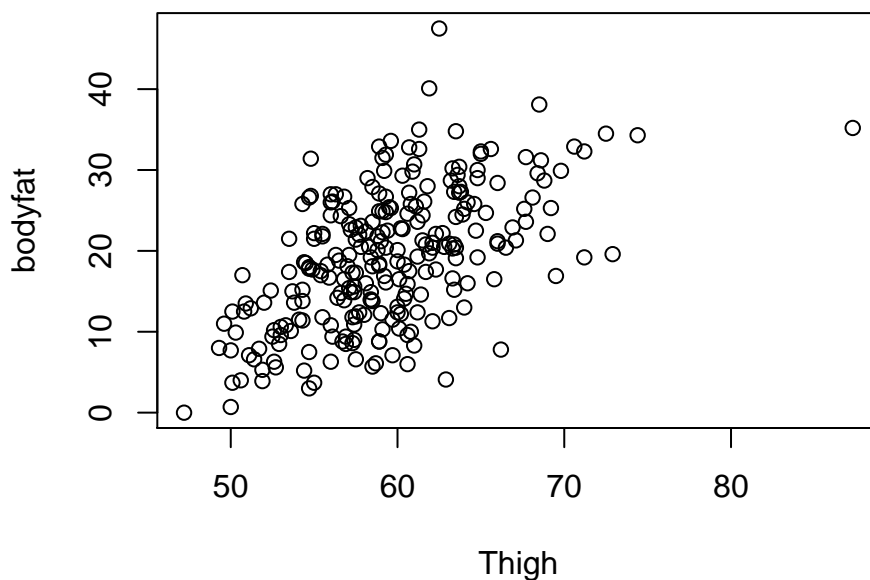
### 3.0.13 13) O modelo está bem ajustado?

```
summary(coef_3)
```

```
##
## Call:
## lm(formula = bodyfat ~ Neck)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2331  -5.3819  -0.2498   5.4919  22.9312
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45.0150     7.2229  -6.232 1.94e-09 ***
## Neck         1.6889     0.1897   8.902 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.307 on 250 degrees of freedom
## Multiple R-squared:  0.2407, Adjusted R-squared:  0.2376
## F-statistic: 79.24 on 1 and 250 DF,  p-value: < 2.2e-16
```

### 3.0.14 14) Faça um gráfico de dispersão entre Bodyfat versus Thigh.

```
plot(bodyfat~Thigh)
```



3.0.15 15) Existe correlação? Qual o tipo de correlação entre as variáveis?

```
cor(bodyfat,Thigh)
```

```
## [1] 0.5596075
```

3.0.16 16) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(bodyfat,Thigh)
```

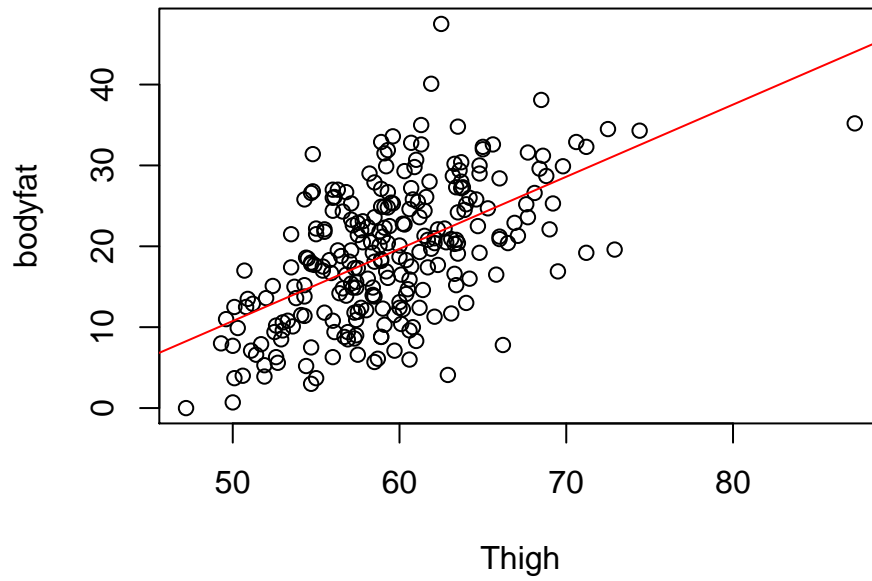
```
##  
## Pearson's product-moment correlation  
##  
## data: bodyfat and Thigh  
## t = 10.676, df = 250, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4684275 0.6389926  
## sample estimates:  
## cor  
## 0.5596075
```

3.0.17 17) Calcule os coeficientes de regressão.

```
coef_4 <- lm(bodyfat ~ Thigh)
```

3.0.18 18) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(bodyfat~Thigh)  
abline(coef_4,col = "red")
```



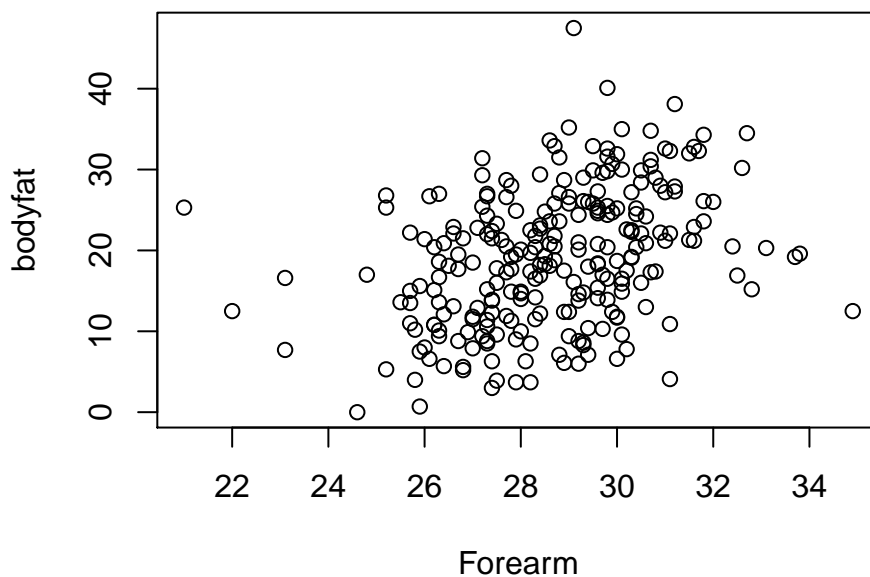
### 3.0.19 19) O modelo está bem ajustado?

```
summary(coef_4)
```

```
##
## Call:
## lm(formula = bodyfat ~ Thigh)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.1677  -4.7837  -0.0614   4.6493  25.5892
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33.84218    4.98282  -6.792 8.03e-11 ***
## Thigh         0.89205    0.08355  10.676 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.95 on 250 degrees of freedom
## Multiple R-squared:  0.3132, Adjusted R-squared:  0.3104
## F-statistic: 114 on 1 and 250 DF, p-value: < 2.2e-16
```

### 3.0.20 20) Faça um gráfico de dispersão entre Bodyfat versus Forearm.

```
plot(bodyfat~Forearm)
```





3.0.21 21) Existe correlação? Qual o tipo de correlação entre as variáveis?

```
cor(bodyfat,Forearm)
```

```
## [1] 0.3613869
```

3.0.22 22) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(bodyfat,Forearm)
```

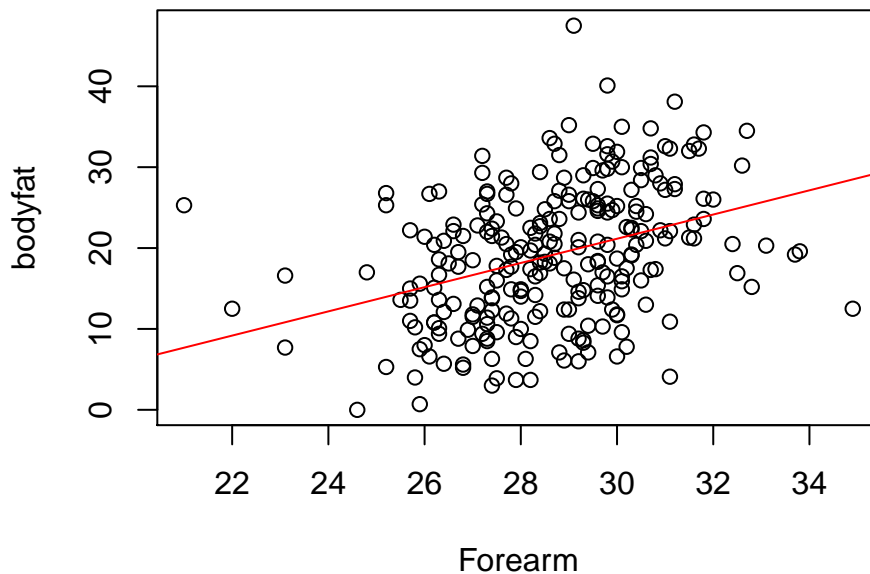
```
##
## Pearson's product-moment correlation
##
## data: bodyfat and Forearm
## t = 6.1282, df = 250, p-value = 3.437e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2489307 0.4642284
## sample estimates:
## cor
## 0.3613869
```

3.0.23 23) Calcule os coeficientes de regressão.

```
coef_5 <- lm(bodyfat ~ Forearm)
```

3.0.24 24) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(bodyfat~Forearm)
abline(coef_5,col = "red")
```



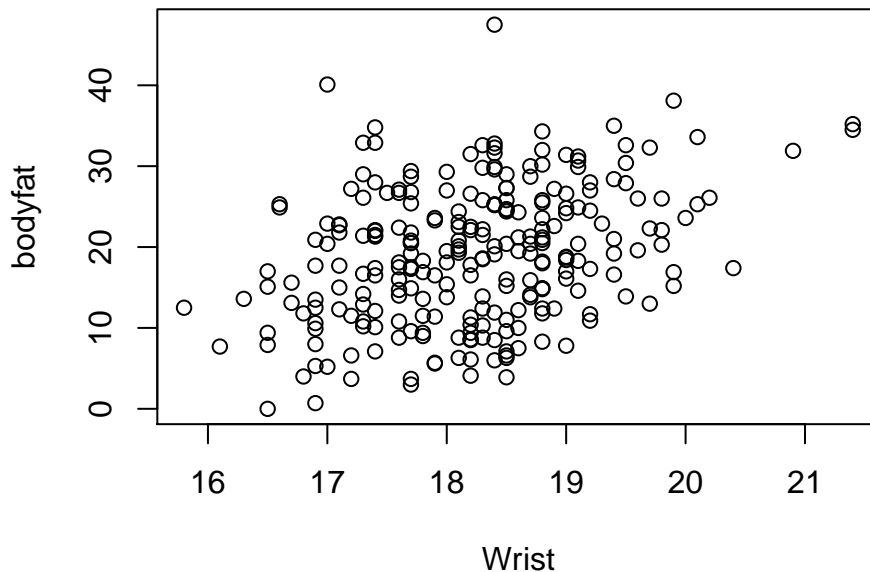
3.0.25 25) O modelo está bem ajustado?

```
summary(coef_5)

##
## Call:
## lm(formula = bodyfat ~ Forearm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.6969  -5.7341  -0.2711   5.6372  27.6965
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -23.7502     7.0179  -3.384 0.000829 ***
## Forearm        1.4967     0.2442   6.128 3.44e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.819 on 250 degrees of freedom
## Multiple R-squared:  0.1306, Adjusted R-squared:  0.1271
## F-statistic: 37.55 on 1 and 250 DF,  p-value: 3.437e-09
```

3.0.26 26) Faça um gráfico de dispersão entre Bodyfat versus Wrist.

```
plot(bodyfat~Wrist)
```



3.0.27 27) Existe correlação? Qual o tipo de correlação entre as variáveis?

```
cor(bodyfat,Wrist)
```

```
## [1] 0.3465749
```

3.0.28 28) Calcule o teste para o coeficiente de correlação. Qual a conclusão do teste?

```
cor.test(bodyfat,Wrist)
```

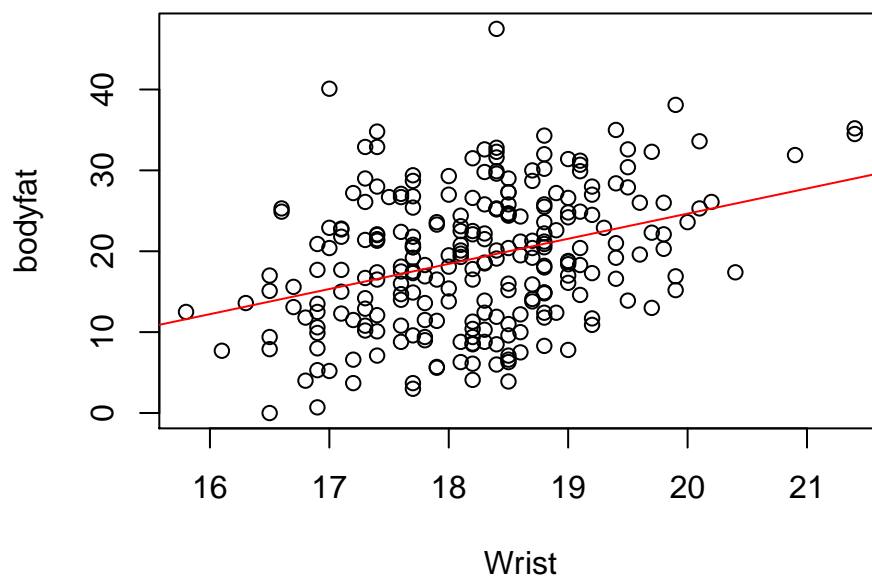
```
##
## Pearson's product-moment correlation
##
## data: bodyfat and Wrist
## t = 5.8419, df = 250, p-value = 1.6e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.2329799 0.4508395
## sample estimates:
## cor
## 0.3465749
```

3.0.29 29) Calcule os coeficientes de regressão.

```
coef_6 <- lm(bodyfat ~ Wrist)
```

3.0.30 30) Acrescente a reta de regressão no gráfico de dispersão.

```
plot(bodyfat~Wrist)  
abline(coef_6,col = "red")
```



### 3.0.31 31) O modelo está bem ajustado?

```
summary(coef_6)
```

```
##
## Call:
## lm(formula = bodyfat ~ Wrist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.0904  -6.0323   0.2917   5.5052  27.8203
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -37.4841     9.7073  -3.861 0.000144 ***
## Wrist        3.1067     0.5318   5.842 1.6e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.866 on 250 degrees of freedom
## Multiple R-squared:  0.1201, Adjusted R-squared:  0.1166
## F-statistic: 34.13 on 1 and 250 DF,  p-value: 1.6e-08
```