

lista 3 - descritiva

Lista 3 - Estatística Descritiva

Professora: Márcia D'Elia Branco

Nomes:

Bruna Umino

Beatriz Vianna

#Exercício 1

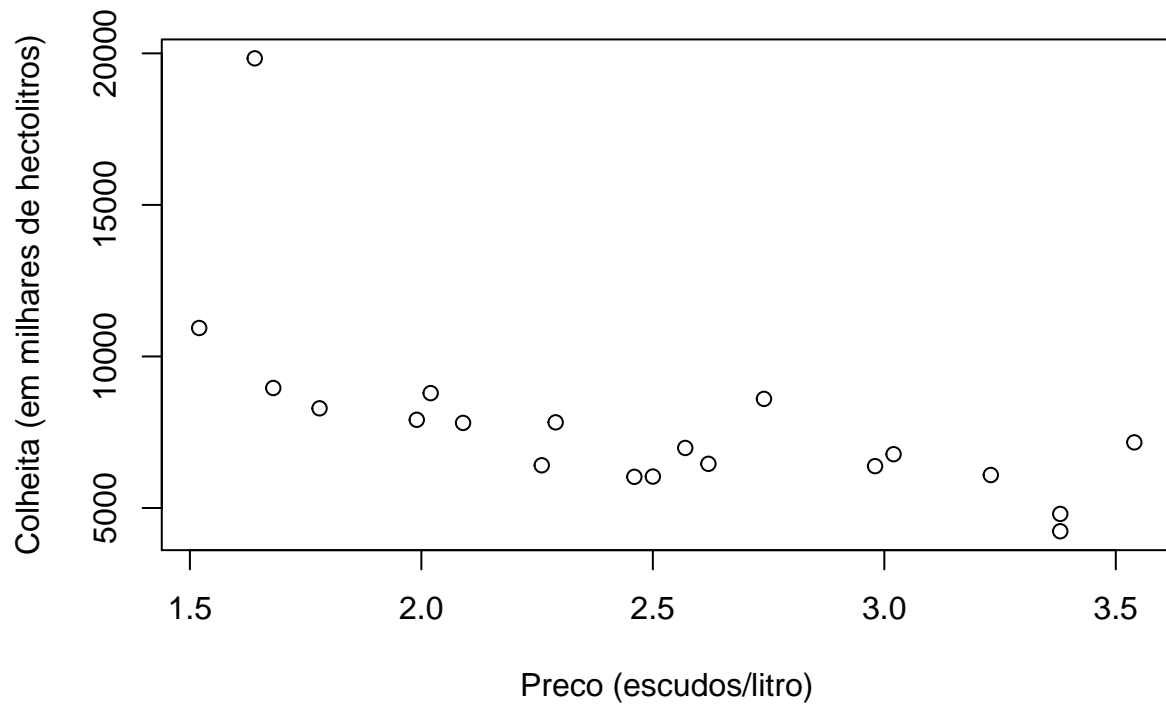
1a

```
ano<-c(1942,1943,1944,1945,1946,1947,1948,1949,1950,1951,
       1952,1953,1954,1955,1956,1957,1958,1959,1960,1961)

colheita<-c(6409, 19835, 10939, 7826, 7165, 7807, 6028, 6037, 6458, 6981,
            4233, 8790, 8959, 8289, 7910, 6775, 6088, 6381, 8600, 4805)

preco<-c(2.26, 1.64, 1.52, 2.29, 3.54, 2.09, 2.46, 2.50, 2.62, 2.57,
         3.38, 2.02, 1.68, 1.78, 1.99, 3.02, 3.23, 2.98, 2.74, 3.38)
tabela <- data.frame(ano, colheita, preco)
#Questao 1a
#Grafico de dispersao
plot( colheita ~ preco, xlab = "Preco (escudos/litro)", ylab = "Colheita (em milhares de hectolitros)",
```

Gráfico de dispersão



```
#Correlacao linear  
cor(preco, colheita)
```

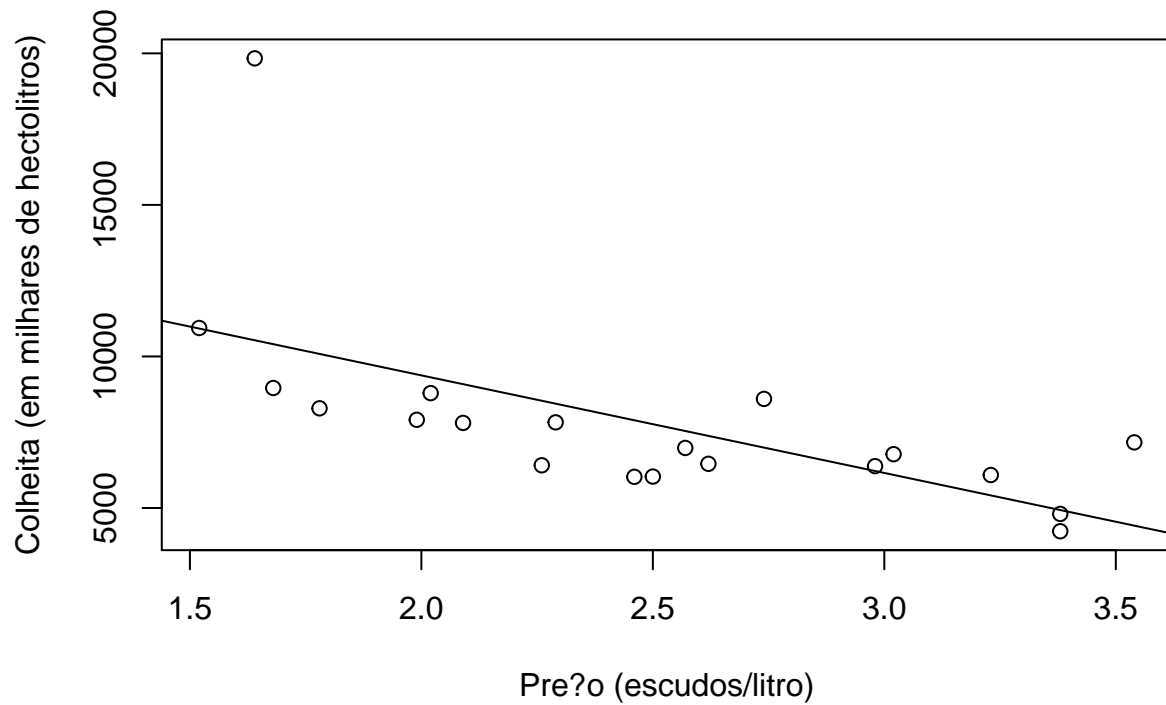
```
## [1] -0.6239457
```

Como podemos observar o gráfico apresenta uma correlação linear significativa e negativa (ou seja, quanto menor a colheita, mais alto fica o preço), que está sendo influenciada pelo dado do ano de 1943, devido a este valor, a correlação está mais elevada.

1b

```
plot( colheita ~ preco, xlab = "Preço (escudos/litro)", ylab = "Colheita (em milhares de hectolitros)",  
abline(lm (colheita ~ preco))
```

Gráfico de dispersão com reta de regressão



```
lm (colheita ~ preco)
```

```
##  
## Call:  
## lm(formula = colheita ~ preco)  
##  
## Coefficients:  
## (Intercept)      preco  
##      15813      -3219
```

Dado que o resultado do coeficiente angular foi igual a -3219, podemos observar que consiste em um valor negativo, ou seja, o preço e a colheita são inversamente proporcionais.

```
#lc
```

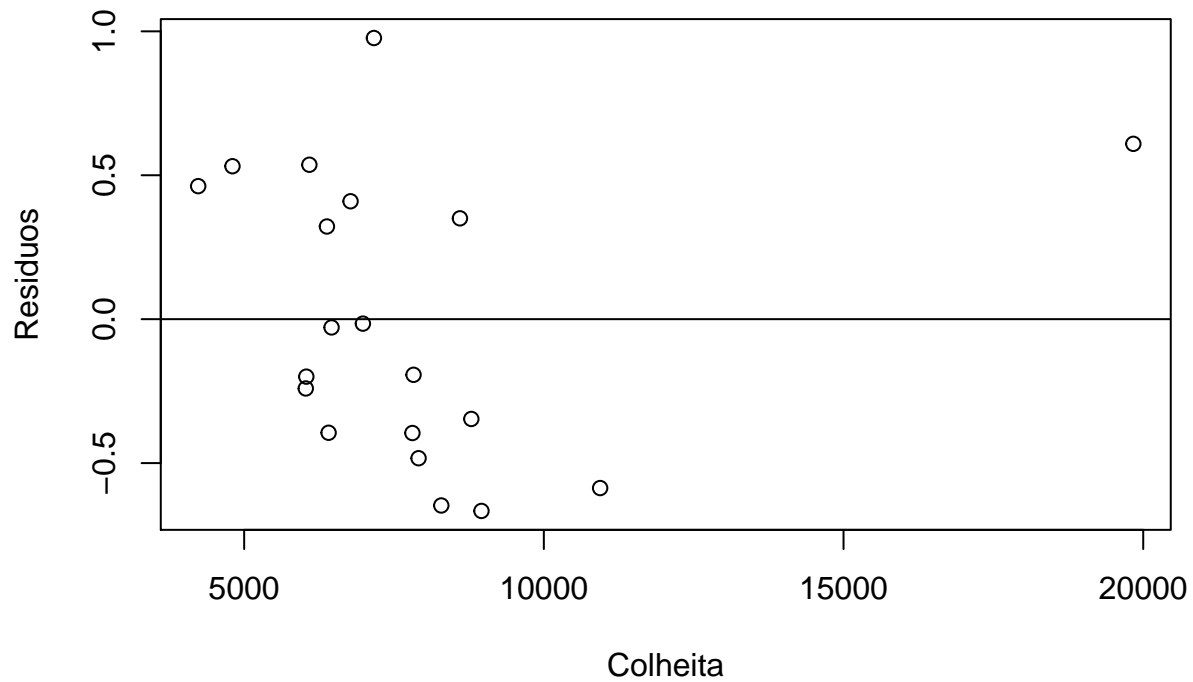
```
lm(preco~colheita)
```

```
##  
## Call:  
## lm(formula = preco ~ colheita)  
##  
## Coefficients:  
## (Intercept)      colheita  
##    3.4297758   -0.0001209
```

```
residuos <- resid(lm(preco~colheita))
```

```
plot(colheita, residuos, ylab="Resíduos", xlab="Colheita", main="Gráfico de resíduos")  
abline(0,0)
```

Gráfico de resíduos



COMENTAR!!!!!!!!!!!!!!

#1d

#diminuir o grafico

`par(mar=c(4,4,4,4))`

#grafico de disperssao ano x colheita

`plot (colheita~ano, type="o", col="darkorchid3", ylab = "", pch=16, main = "Gráfico colheita x ano x pr`

#legendas

`mtext("Colheita", side = 2, line = 2.5, col="darkorchid3")`

`par(new=TRUE)`

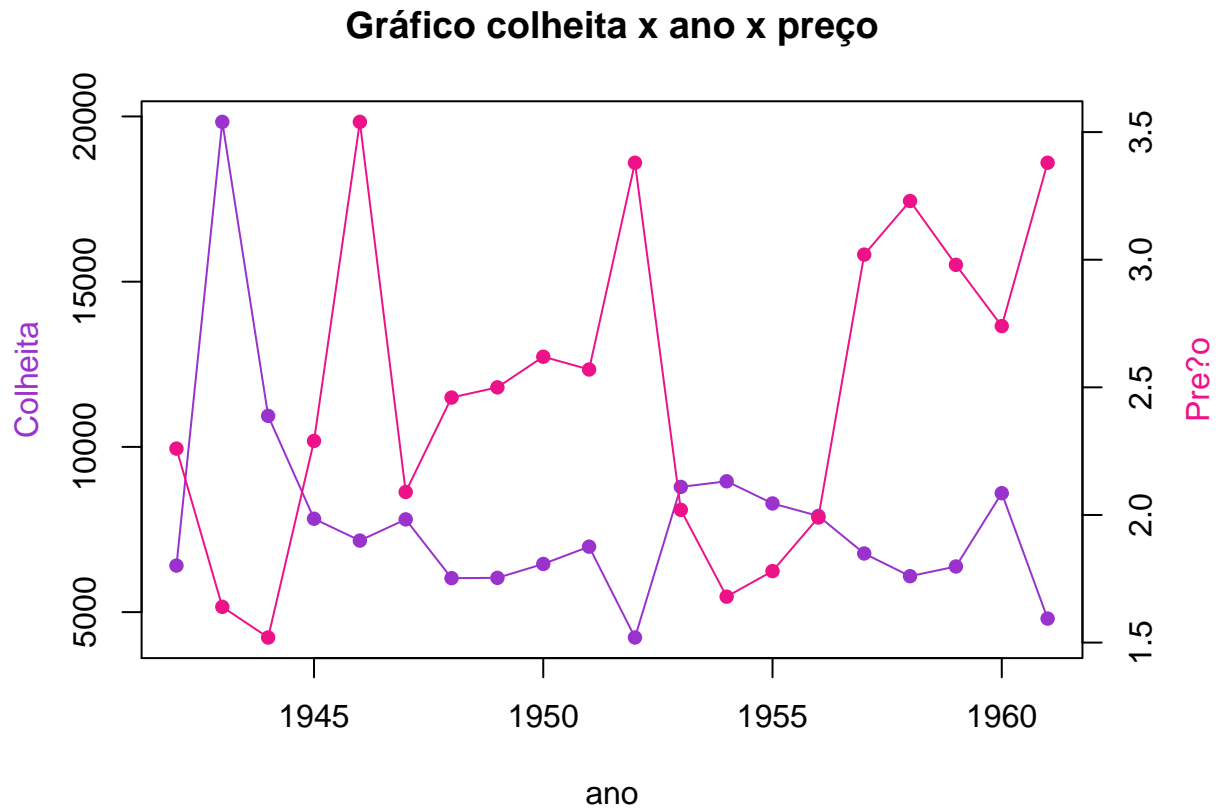
#grafico de disperssao ano x pre?o

`plot(preco~ano, axes=FALSE, type='o', col='deeppink2', ann=FALSE, pch=16)`

#legendas

`mtext("Pre?o", side = 4, line = 2.5, col="deeppink2")`

`axis(4)`

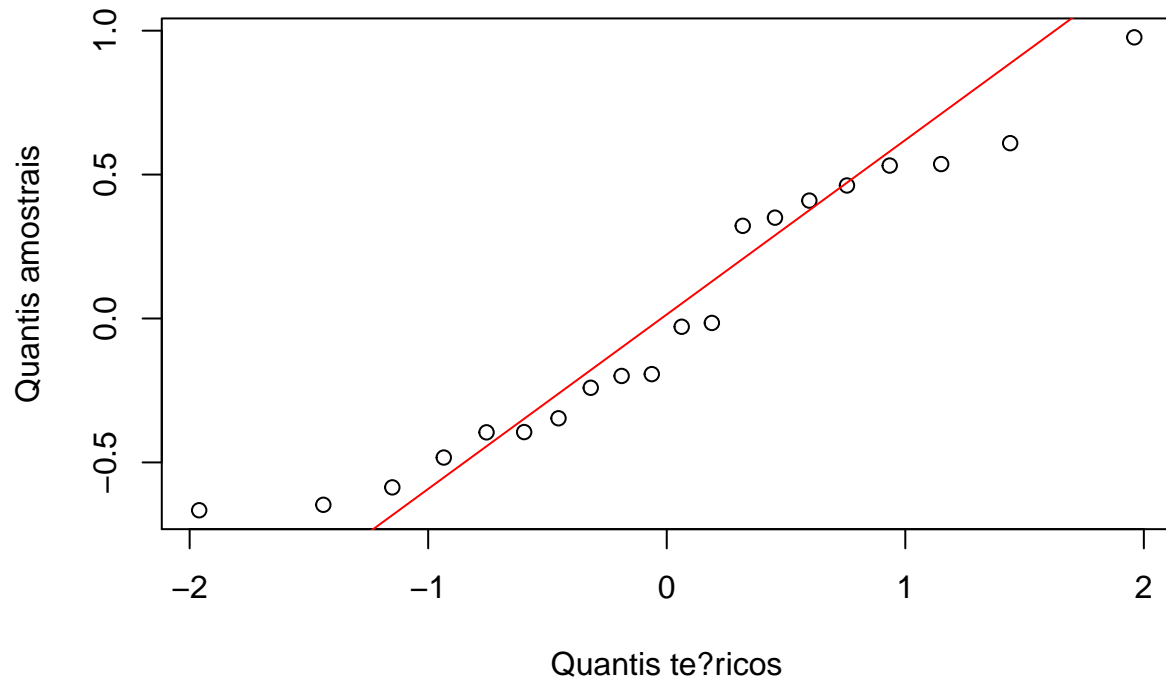


COMENTAR!!!!!!!!!!!!!!!!!!!!!! Através dos gráficos podemos observar que ocorreu uma considerável colheita no ano de 1943, que resultou no pior preço e no ano de 1952 ocorreu a menor colheita deste intervalo de tempo, o resto dos valores coletados estão na faixa de 5000 a 10900 milhares de hectolitros. Em relação ao preço há picos nos anos de 1946, 1961, 1952 e 1958, que são os anos os quais ocorreram as piores colheitas e nos anos de 1944, 1943 e 1954, foi relatado os menores preços.

1e

```
qqnorm(residuos, main = "Gráfico de probabilidades normais", xlab = "Quantis teóricos", ylab = "Quantis amostrais")
qqline(residuos, col="red")
```

Gráfico de probabilidades normais



ANALISAR!!!!!!!!!!!!!!!!!!!!

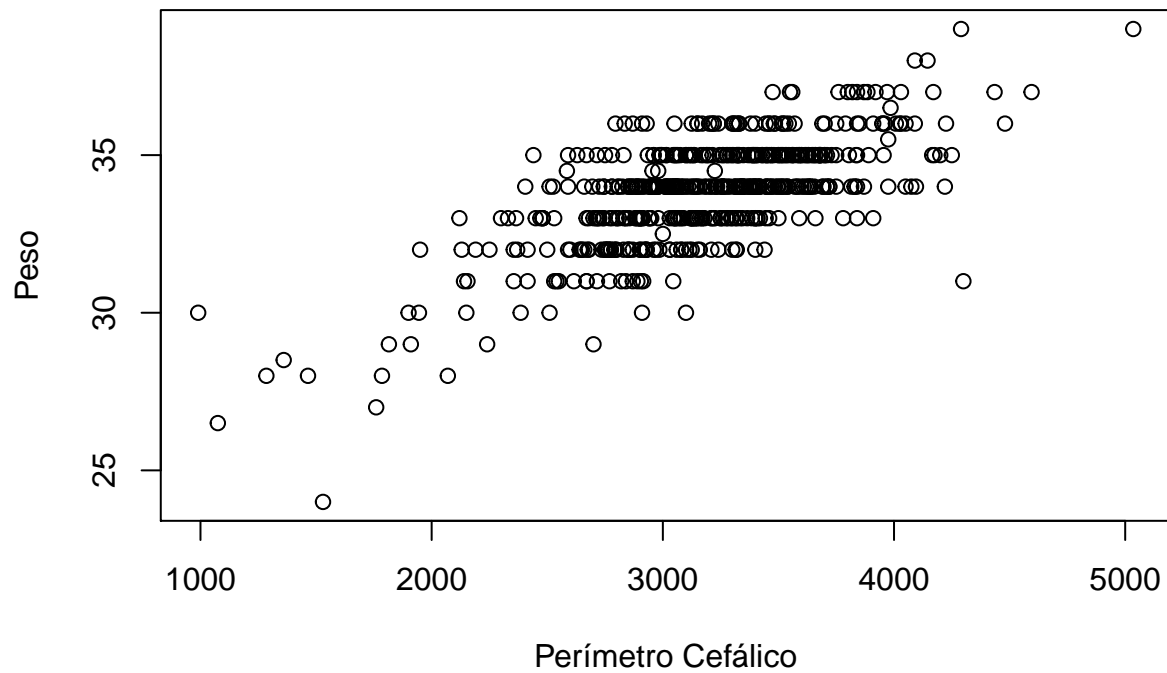
Questao 2

2a

```
library (magrittr)
dados <- read.csv2("/home/be/viabianna/Downloads/dadosmalariaCEA15P14.csv")
#Retirar os dados que contém N/A
dados %>% subset(!is.na(pc)) %>% subset(!is.na(peso)) %>% subset(!is.na(est))

#Gráfico de Dispersão Perímetro Cefálico x Peso
plot(dados$pc~dados$peso, xlab="Perímetro Cefálico", ylab="Peso", main= "Gráfico de Dispersão Perímetro Cefálico x Peso")
```

Gráfico de Dispersão Perímetro Cefálico x Peso



```
#Correlação Perímetro Cefálico x Peso
```

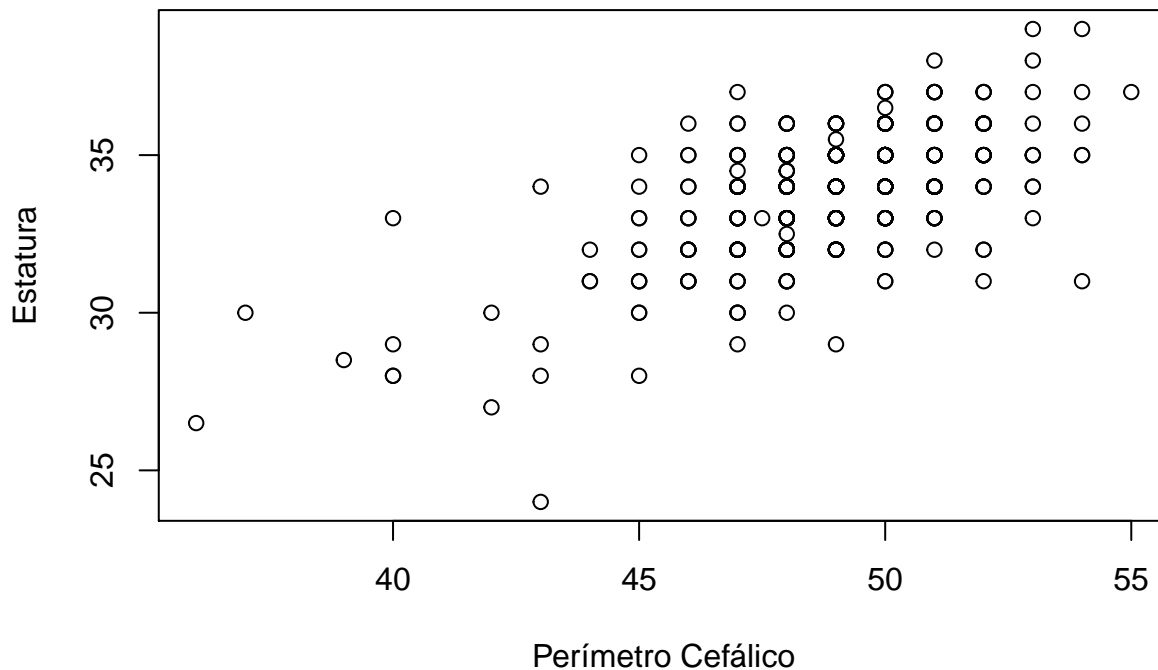
```
cor(dados$pc,dados$peso)
```

```
## [1] NA
```

```
#Gráfico de Dispersão Perímetro Cefálico x Estatura
```

```
plot(dados$pc~dados$est, xlab="Perímetro Cefálico", ylab="Estatura", main= "Gráfico de Dispersão Perímetro Cefálico x Estatura")
```

Gráfico de Dispersão Perímetro Cefálico x Estatura



```
#Correlação Perímetro Cefálico x Estatura
cor(dados$pc,dados$est)
```

```
## [1] NA
```

```
#2b
```

```
equação <- (lm(dados$pc~dados$peso))
equação
```

```
##
## Call:
## lm(formula = dados$pc ~ dados$peso)
##
## Coefficients:
## (Intercept)  dados$peso
##    26.06791    0.00244
```

A partir destes dados, sabemos que a reta de regressão para Perímetro Cefálico x Peso (que é a variável que apresenta maior correlação) será

$$y = 0.002x + 26.061$$

portanto, o perímetro cefálico, em centímetros, esperado para um bebê de 3000g é:

$$y = 0.002 * 3000 + 26.061$$

$$y = 32.061$$

```
equação <- (lm(dados$pc~dados$est))
equação
```

```
##
## Call:
```



```
## lm(formula = dados$pc ~ dados$est)
##
## Coefficients:
## (Intercept)      dados$est
##      10.2103      0.4824
```

Já a reta Perímetro Cefálico x Estatura (que apresenta menor correlação) será

$$y = 0.482x + 10.21$$

Assim sendo, o perímetro cefálico esperado, em centímetros, para um recém nascido de 50cm será:

$$y = 0.482 * 50 + 10.21$$

$$y = 34.31 \text{ #2c}$$

```
dados2 <- data.frame(dados$peso, dados$grupo)
dados2 <- subset(dados2, !is.na(dados$peso))
dadosgrupo <- vector(length=length(dados2$dados.grupo))
dadosgrupo[which(dados2$dados.grupo==1)] <- 1
dadosgrupo[which(dados2$dados.grupo==2)] <- 1
dadosgrupo[which(dados2$dados.grupo==3)] <- 1
dados2$dados.grupo <- dadosgrupo
```