MAE0217 - Estatística Descritiva - Lista 4

Natalia Hitomi Koza[1]
Rafael Gonçalves Pereira da Silva[2]
Ricardo Geraldes Tolesano[3]
Rubens Kushimizo Rodrigues Xavier[4]
Rubens Gomes Neto[5]
Rubens Santos Andrade Filho[6]
Thamires dos Santos Matos[7]

Junho de 2021

Sumário

[1]Número USP: 10698432
[2]Número USP: 9009600
[3]Número USP: 10734557
[4]Número USP: 8626718
[5]Número USP: 9318484
[6]Número USP: 10370336
[7]Número USP: 9402940

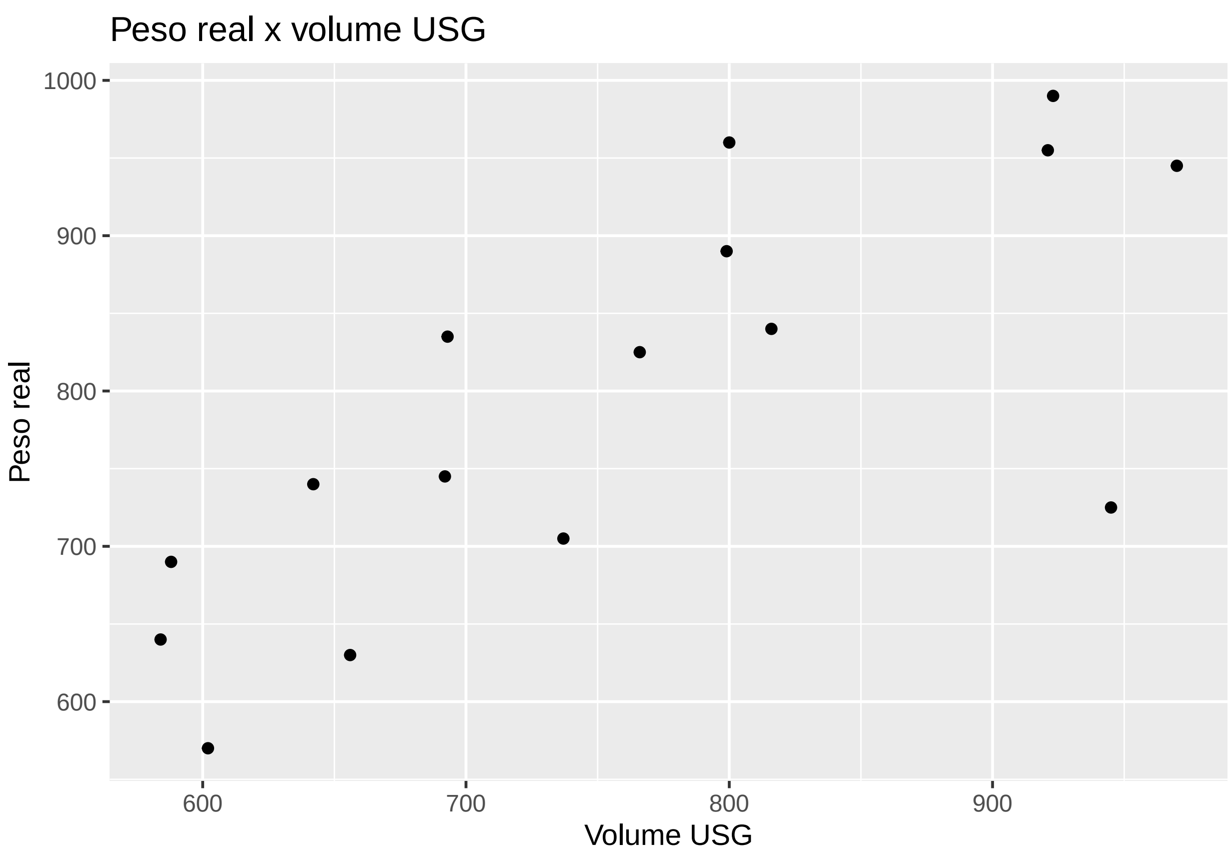
Exercício 1

i)

Tomaremos Volume USG como a variável explicativa x e Peso Real como a variável resposta y. Adotaremos o modelo de regressão linear simples $y_i = \alpha + \beta x_i + e_i$, onde α é o intercepto, *beta* é a inclinação da reta, e e_i são erros aleatórias não correlacionados.

ii)

```
scatter_title <- "Peso real x volume USG"
scatter_x <- "Volume USG"
scatter_y <- "Peso real"
fit_titles <- list("Resíduos vs observações x para o ajuste feito no modelo",
                   "Gráfico Q-Q normal para o ajuste feito no modelo",
                   "Resíduos normalizados vs observações x para o ajuste feito no modelo",
                   "Resíduos normalizados vs influência das observações para o ajuste feito no modelo")


dados1 <- read_excel("data/peso_volume_figado.xlsx")
dados1 <- dados1[order(dados1$volume_usg), ]
# ggplot(dados, aes(x=volume_usg, y=peso_real)) + geom_point() + geom_smooth(method=lm)
questao_i <- function(dados) {}
ggplot(dados1, aes(x=volume_usg, y=peso_real)) + geom_point() + labs(title=scatter_title, x=scatter_x, y
```

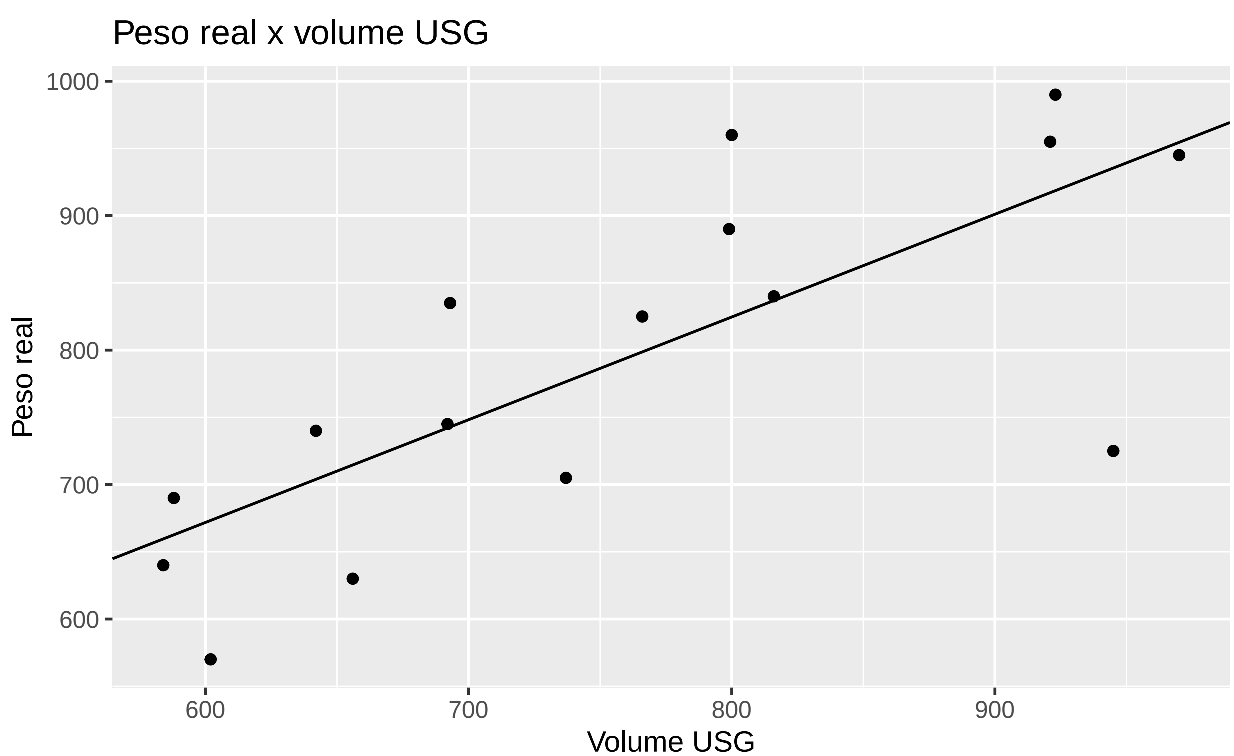


iii)

Realizaremos o ajuste do modelo e mostraremos algumas métricas de qualidade do modelo:

```
ajustarModelo <- function(dados) {
  ajuste <- lm(peso_real ~ volume_usg, data=dados)
  intercept <- ajuste$coefficients[1]
  slope <- ajuste$coefficients[2]
  print("O ajuste encontrou os coeficientes:")
  print(paste("Alpha:", intercept))
  print(paste("Beta:", slope))
  p <- ggplot(dados, aes(x=volume_usg, y=peso_real)) + geom_point() + geom_abline(intercept = intercept
  plot(p)
  print(summary(ajuste))
  plot(ajuste,
       caption=fit_titles)

  return(ajuste)
}

ajuste <- ajustarModelo(dados1)
```

```
## [1] "O ajuste encontrou os coeficientes:"
## [1] "Alpha: 213.276155355598"
## [1] "Beta: 0.764181763170465"
```




```
##
```

```
## Call:
## lm(formula = peso_real ~ volume_usg, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -210.43  -32.54   14.76   44.97  135.38
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 213.2762   133.3334   1.600 0.132011
## volume_usg    0.7642     0.1734   4.407 0.000597 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 87.91 on 14 degrees of freedom
## Multiple R-squared:  0.5811, Adjusted R-squared:  0.5512
## F-statistic: 19.42 on 1 and 14 DF,  p-value: 0.000597
```

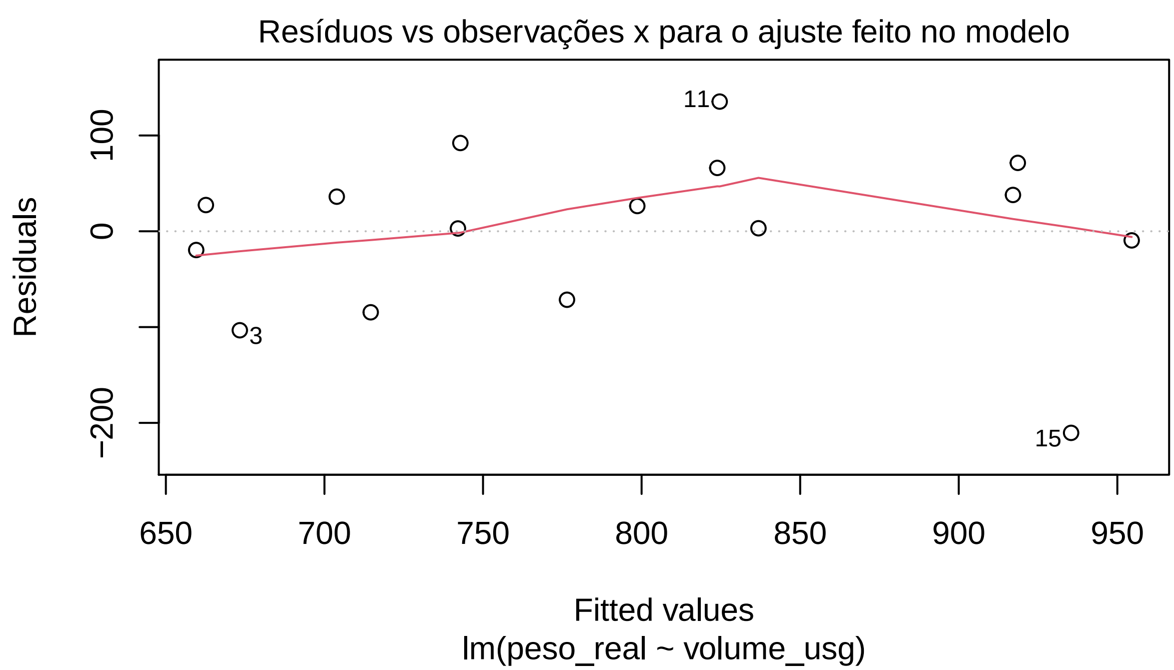
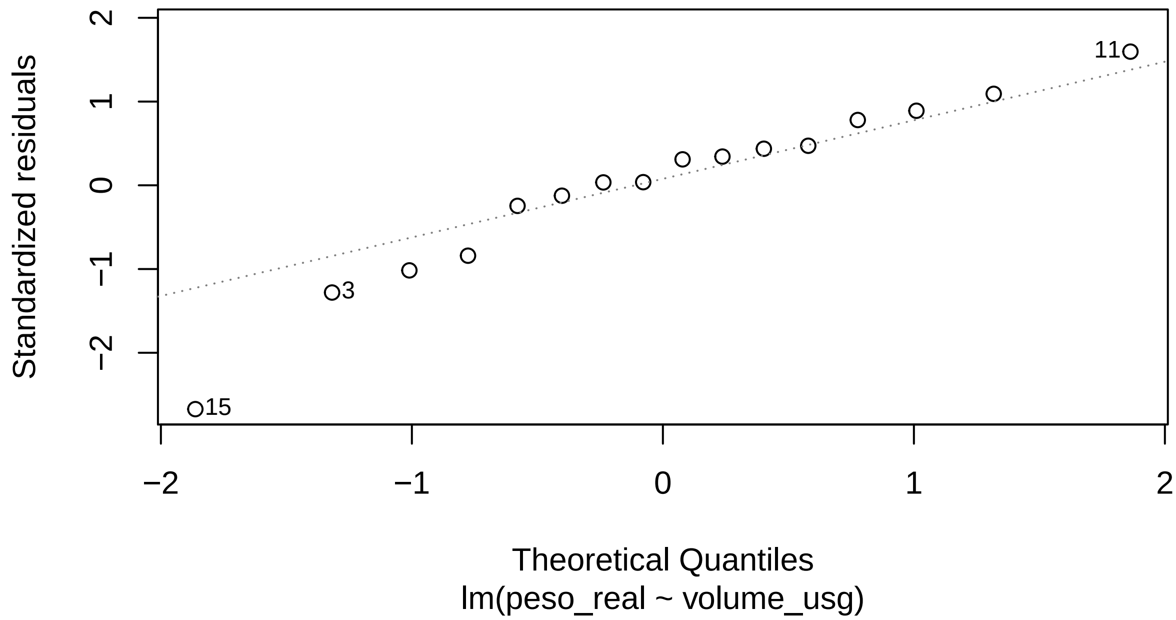
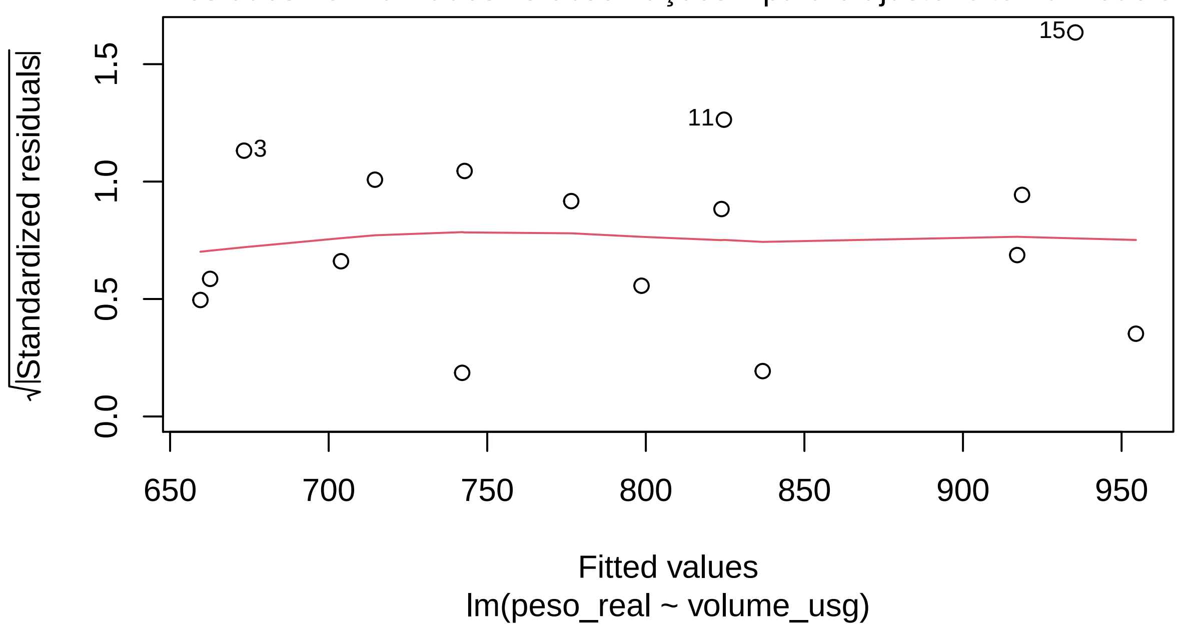
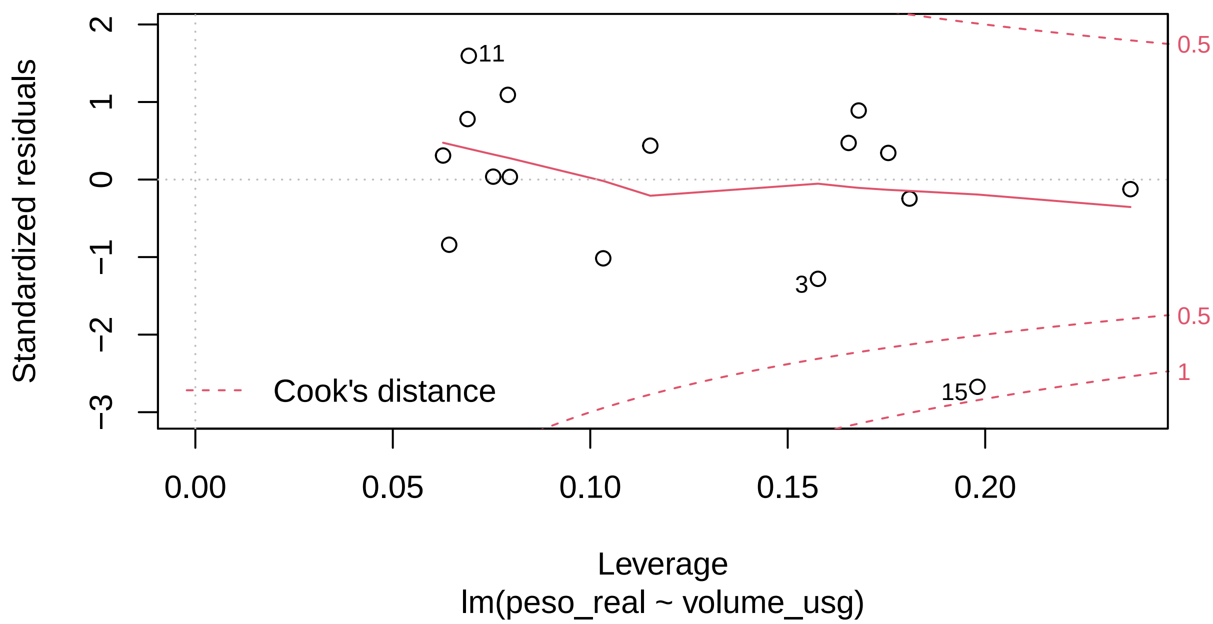


Gráfico Q−Q normal para o ajuste feito no modelo



Resíduos normalizados vs observações x para o ajuste feito no modelo

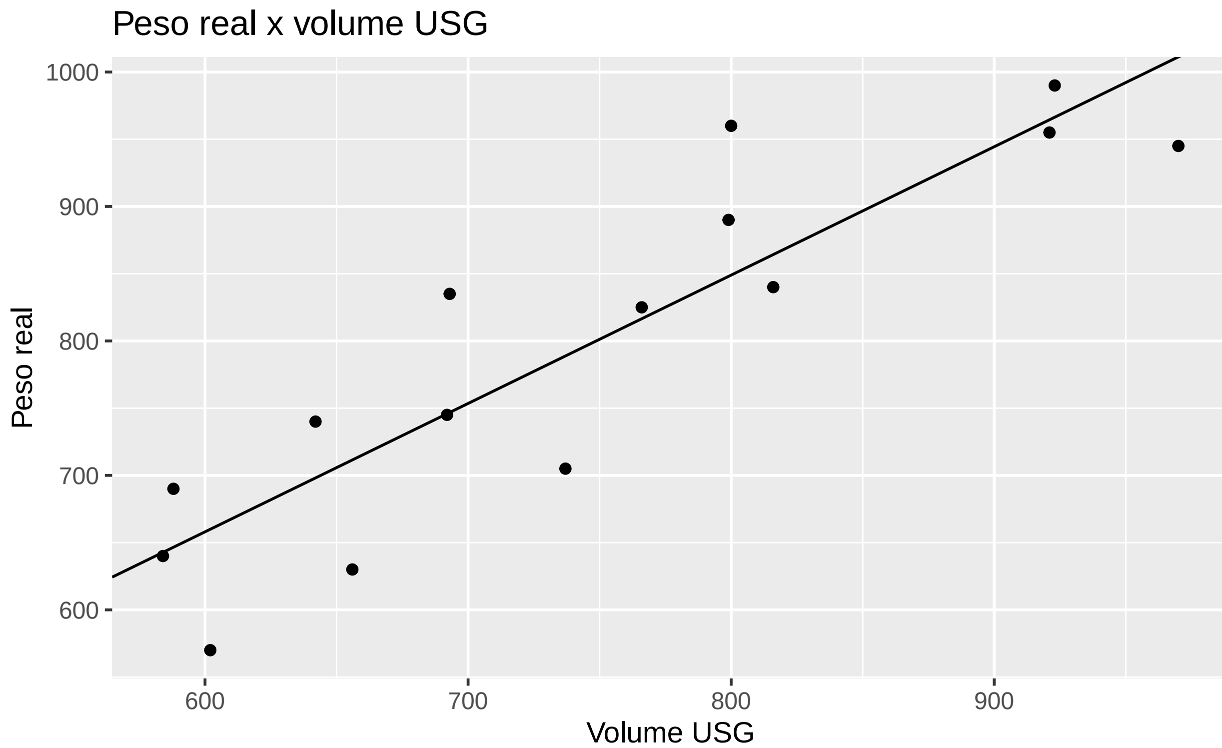




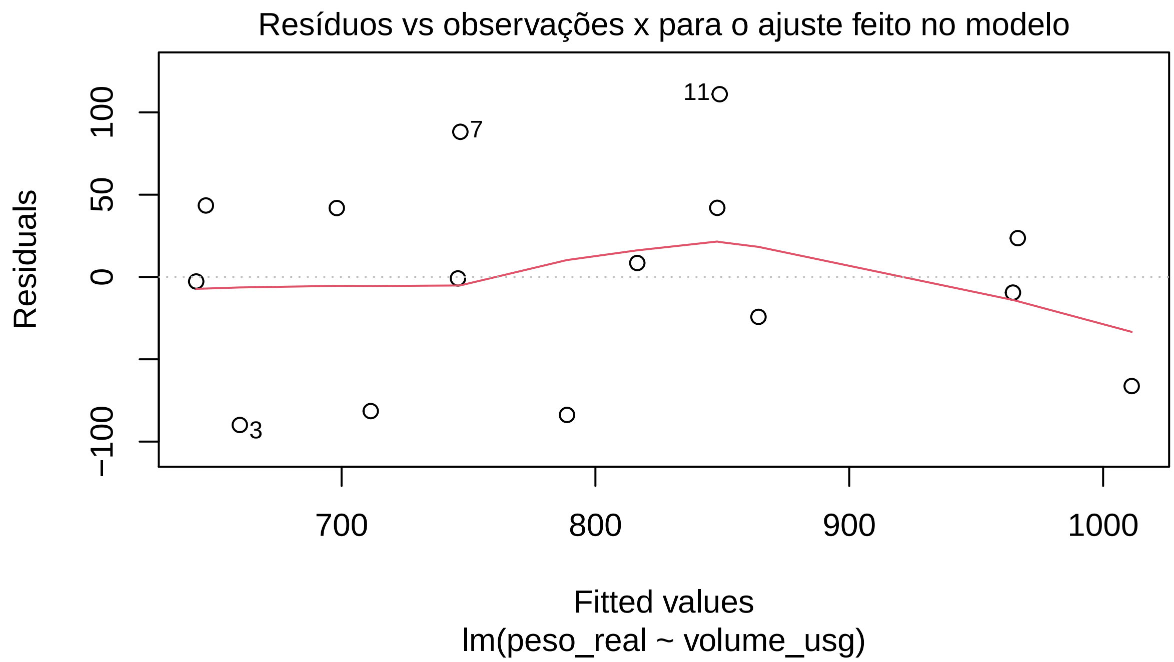
A análise do ajuste indicou que as observações 3, 11 e 15 são mais influentes no modelo. Em especial, a observação 15 se destaca como outlier em todos os gráficos mostrados. Realizaremos novamente o ajuste com essa observação removida. Não removeremos as observações 3 e 11 dado que possuímos poucas observações e elas não fogem do padrão na mesma intensidade elevada da observação 15.

```
dados2 <- dados1[-c(15), ]
ajuste <- ajustarModelo(dados2)
```

```
## [1] "O ajuste encontrou os coeficientes:"
## [1] "Alpha: 85.159261447348"
## [1] "Beta: 0.954742253846616"
```




```
##
## Call:
## lm(formula = peso_real ~ volume_usg, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.914 -45.244  -0.841  41.949 111.047
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  85.1593   102.8848   0.828    0.423
## volume_usg    0.9547     0.1361   7.013 9.17e-06 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.83 on 13 degrees of freedom
## Multiple R-squared:  0.7909, Adjusted R-squared:  0.7748
## F-statistic: 49.18 on 1 and 13 DF,  p-value: 9.167e-06
```


Resíduos vs observações x para o ajuste feito no modelo
Residuals
100
50
0
−100
700
800
900
1000
Fitted values
lm(peso_real ~ volume_usg)
11
7
3

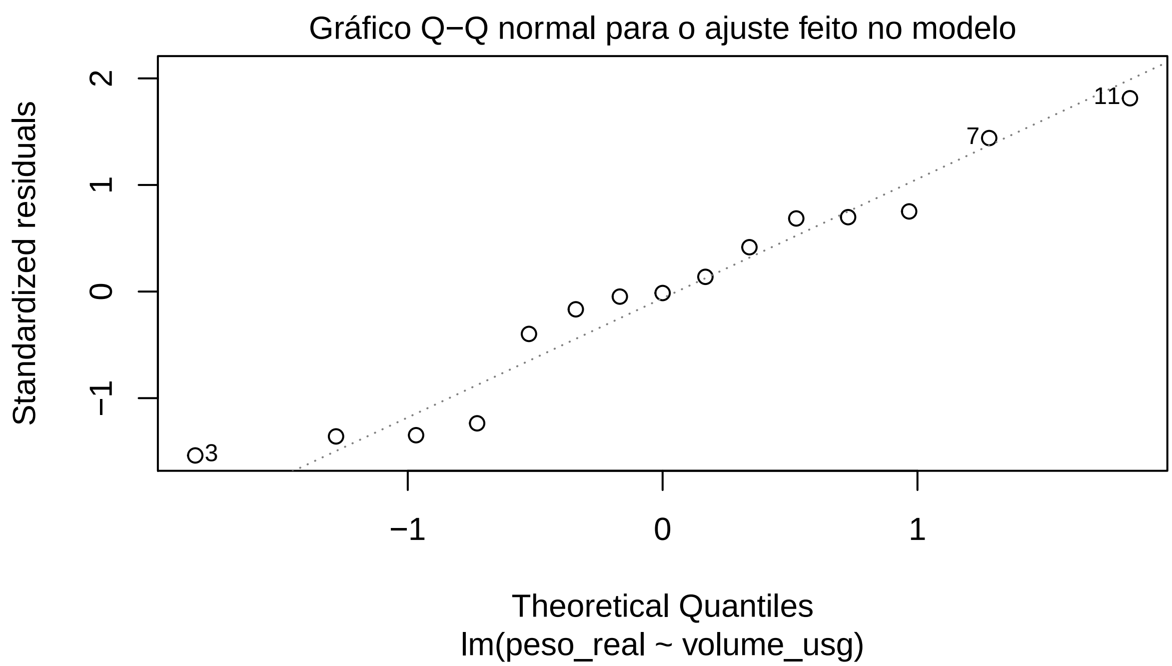
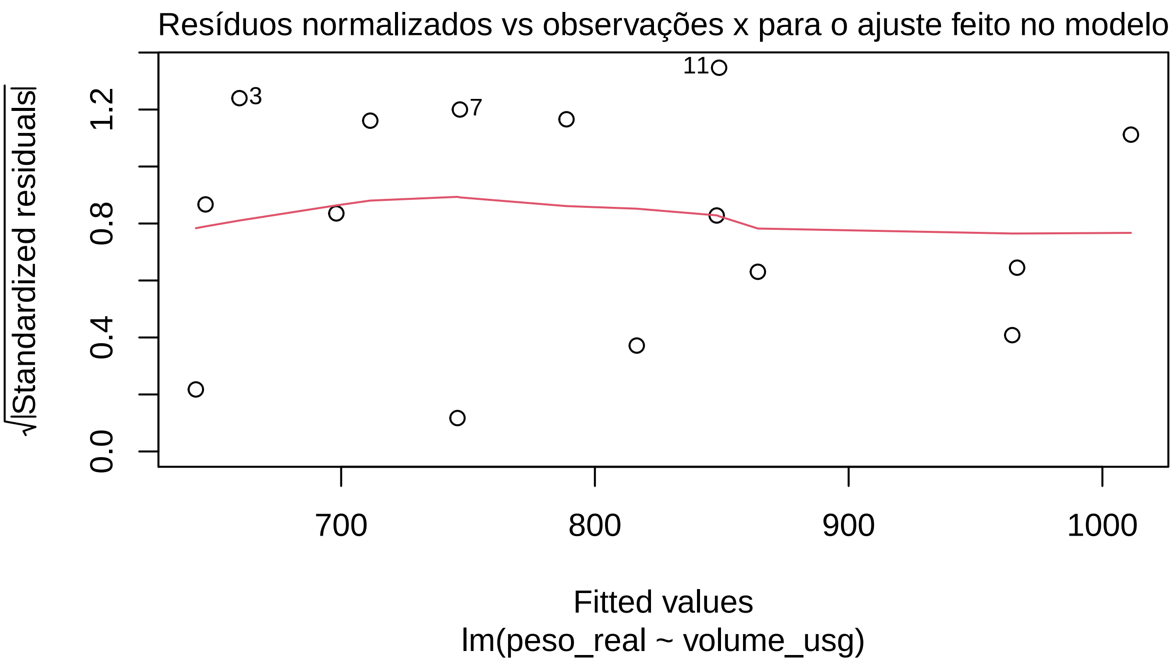
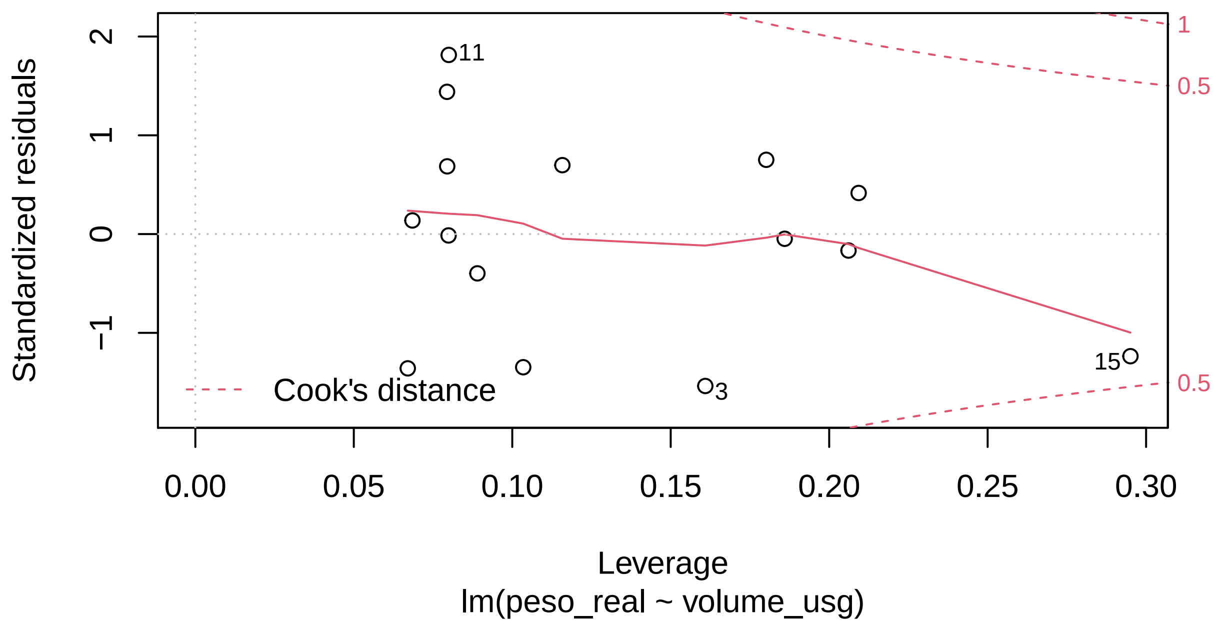


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
2
1
0
−1
−1
0
1
Theoretical Quantiles
lm(peso_real ~ volume_usg)
11
7
3





Observamos uma melhora significativa no valor Rˆ2 após a remoção da observação 15. Os gráficos indicam que os resíduos possuem os valores dentro do esperado. Idealmente, o Rˆ2 deveria estar próximo de 1, mas não está. Dessa forma, podemos concluir que o ajuste do modelo aproxima os dados, mas não estritamente. Assim, espera-se que o intervalo de confiança ao prever o peso real com base no volume seja grande.

iv)

Construindo intervalos de confiança dos parâmetros:

```
confidence_intervals <- confint(ajuste)
rownames(confidence_intervals) <- c("Alpha", "Beta")
kable(confidence_intervals, caption="Intervalos de confiança para o ajuste dos parâmetros do modelo")
```

Tabela 1: Intervalos de confiança para o ajuste dos parâmetros do modelo

	2.5 %	97.5 %
Alpha	-137,11	307,43
Beta	0,66	1,25

v)

A seguir, construiremos a tabela.

```
volumes <- c(600, 700, 800, 900, 1000)
df <- data.frame(volume_usg = volumes)
previsto <- predict(ajuste, df, interval='confidence')
previsto <- data.frame(previsto)
intervalo <- previsto$fit - previsto$lwr
previsto <- cbind(volume_usg = volumes, peso = previsto$fit, intervalo = intervalo)
colnames(previsto) <- c("Volume", "Peso previsto", "Intervalo de confiança de 95%")
kable(previsto, caption="Pesos previstos pelo modelo")
```

Tabela 2: Pesos previstos pelo modelo

Volume	Peso previsto	Intervalo de confiança de 95%
600	658,00	55,77
700	753,48	38,08
800	848,95	39,00
900	944,43	57,63
1.000	1.039,90	82,78

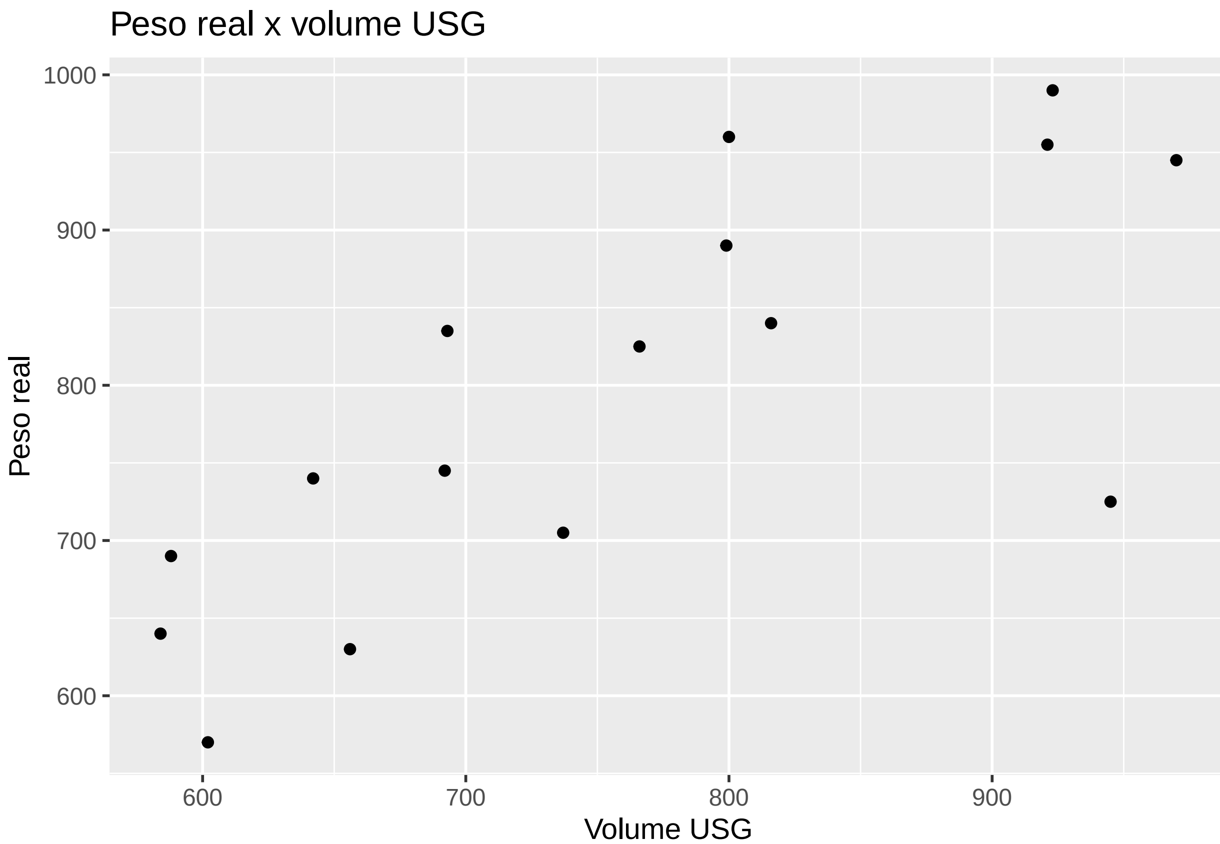
vi)

vi)i)

Novamente, tomaremos o Volume USG como a variável explicativa x e o Peso Real como a variável resposta y. Adotaremos o modelo de regressão linear simples $y_i = \beta x_i + e_i$, onde *beta* é a inclinação da reta e e_i são erros aleatórias não correlacionados.

vi)ii)

```r
dados3 <- data.frame(dados1)
ggplot(dados3, aes(x=volume_usg, y=peso_real)) + geom_point() + labs(title=scatter_title, x=scatter_x, y
```



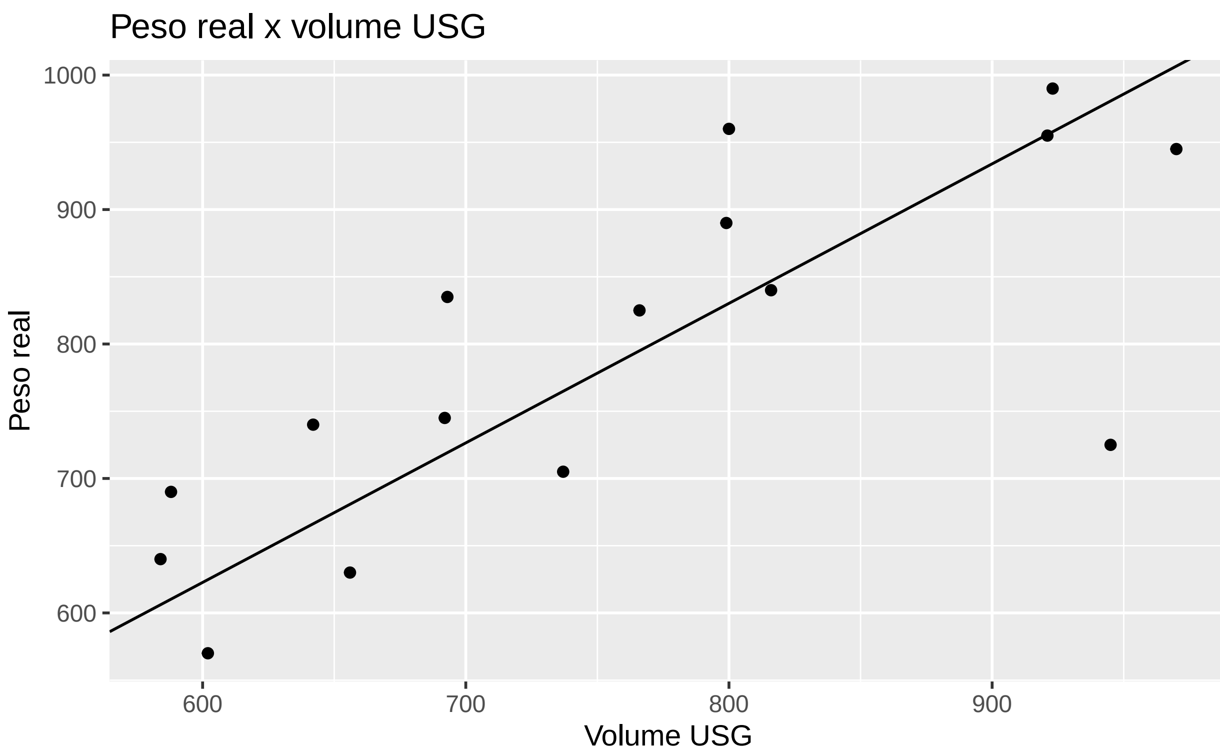
vi)iii)

Realizaremos o ajuste do modelo e mostraremos algumas métricas de qualidade do modelo:

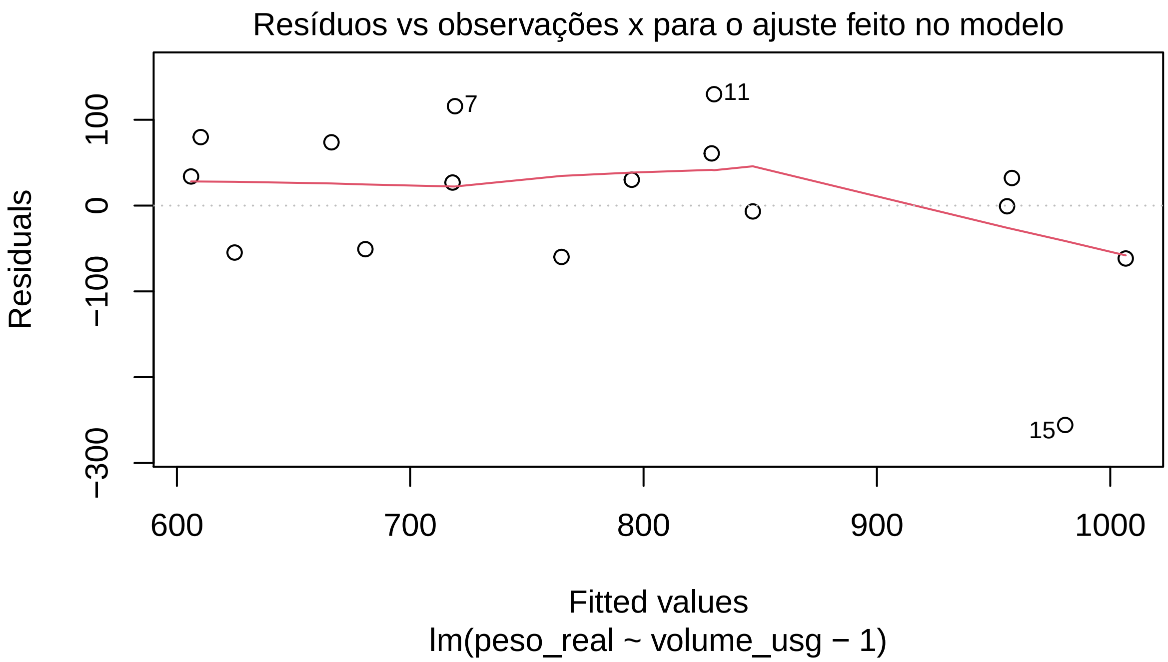
```r
ajustarModelo <- function(dados) {
  # - 1 omite o intercepto
  ajuste <- lm(peso_real ~ volume_usg - 1, data=dados)
  intercept <- 0
  slope <- ajuste$coefficients
  print("O ajuste encontrou o coeficiente:")
  print(paste("Beta:", slope))
  p <- ggplot(dados, aes(x=volume_usg, y=peso_real)) + geom_point() + geom_abline(intercept = intercept
  plot(p)
  print(summary(ajuste))
  plot(ajuste, caption=fit_titles)

  return(ajuste)
}

ajuste <- ajustarModelo(dados3)
```

```
## [1] "O ajuste encontrou o coeficiente:"
```

```
## [1] "Beta: 1.03776957920071"
```




```
##
## Call:
## lm(formula = peso_real ~ volume_usg - 1, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -255.69  -51.77   28.47   64.06  129.78
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## volume_usg  1.03777    0.03003   34.56 1.03e-15 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.36 on 15 degrees of freedom
## Multiple R-squared:  0.9876, Adjusted R-squared:  0.9868
## F-statistic:  1194 on 1 and 15 DF,  p-value: 1.026e-15
```


Resíduos vs observações x para o ajuste feito no modelo
Residuals
100
0
−100
−300
7
11
15
600
700
800
900
1000
Fitted values
lm(peso_real ~ volume_usg − 1)

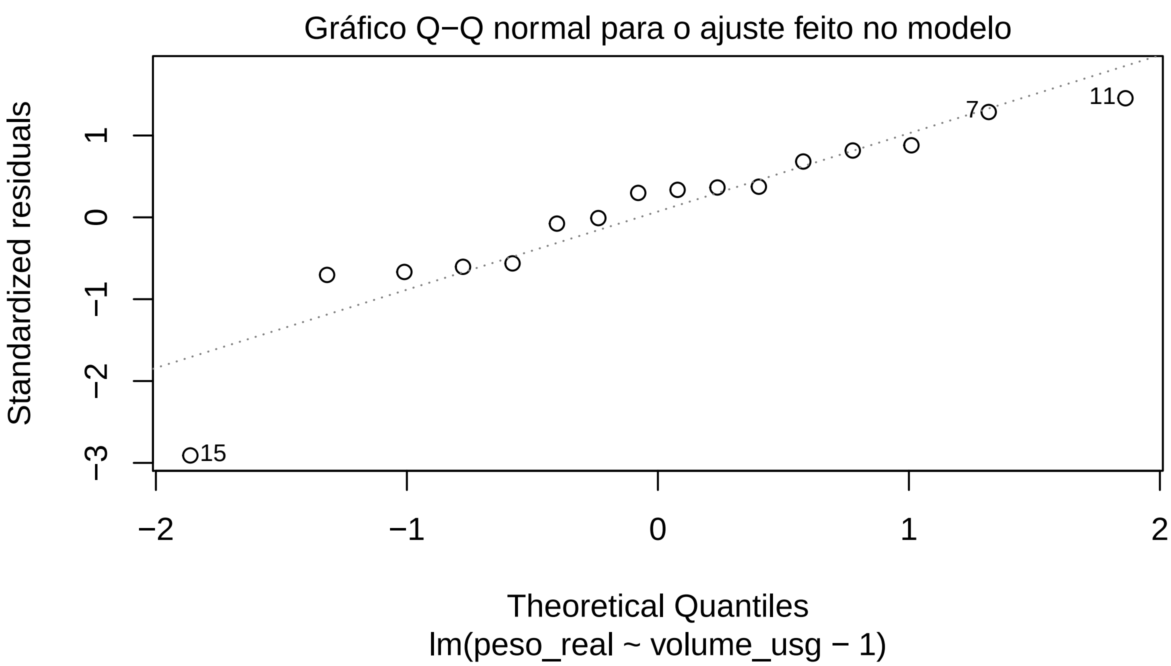
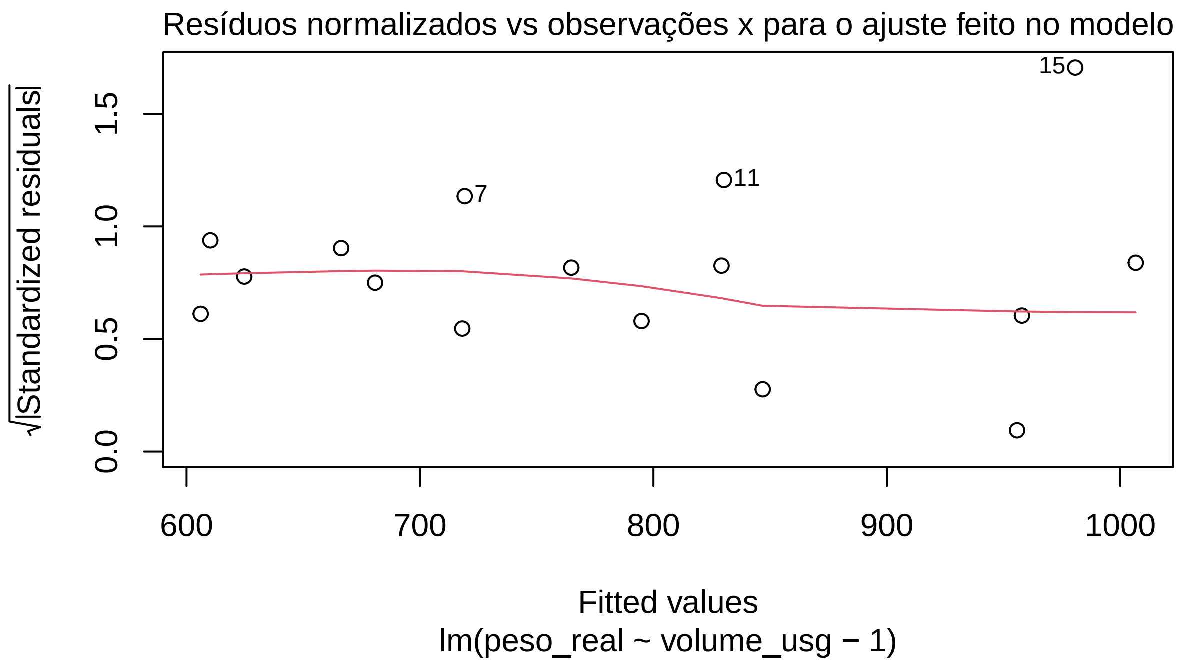
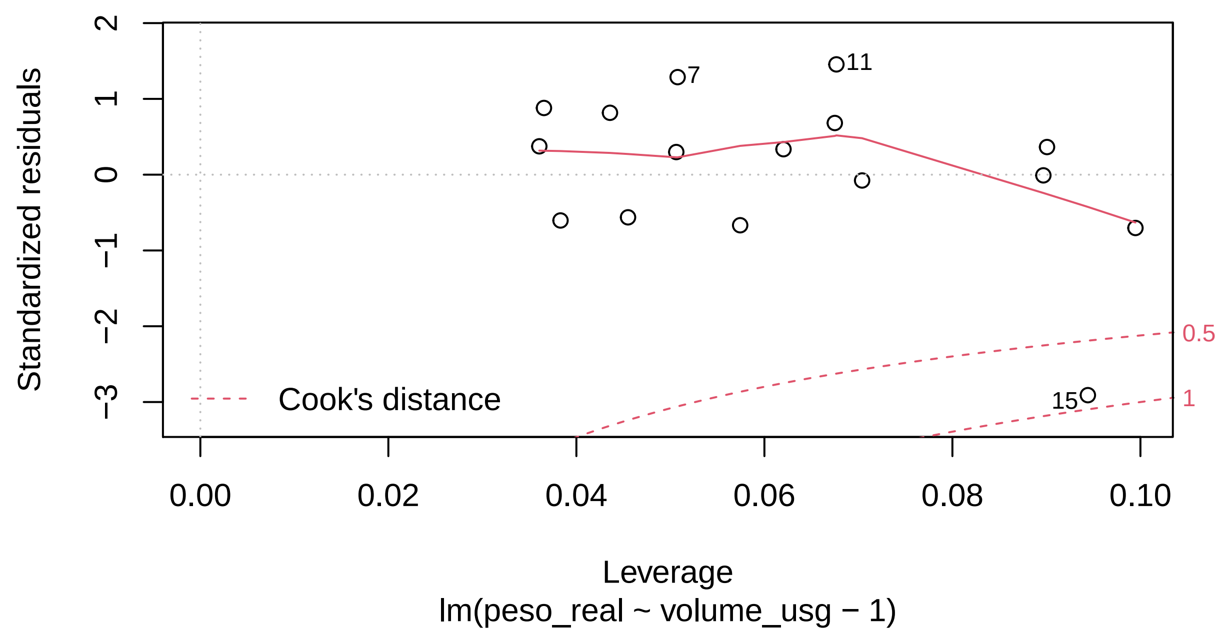


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
1
0
−1
−2
−3
7
11
15
−2
−1
0
1
2
Theoretical Quantiles
lm(peso_real ~ volume_usg − 1)

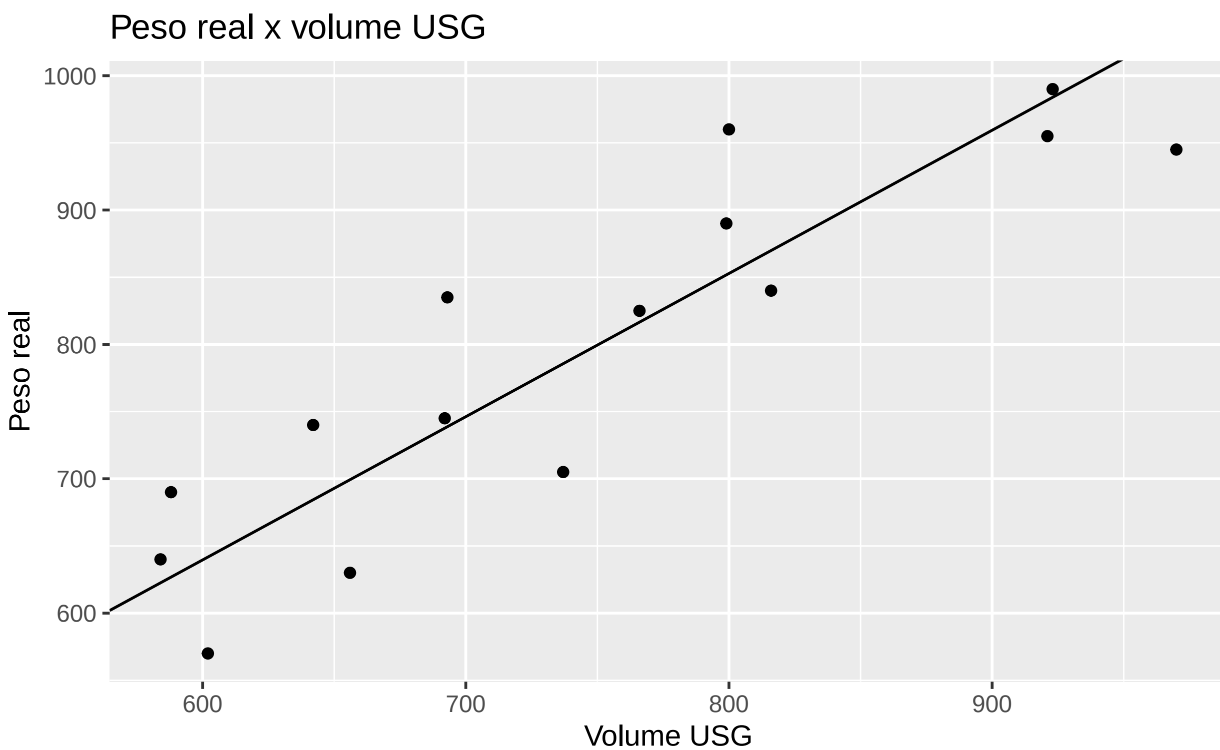




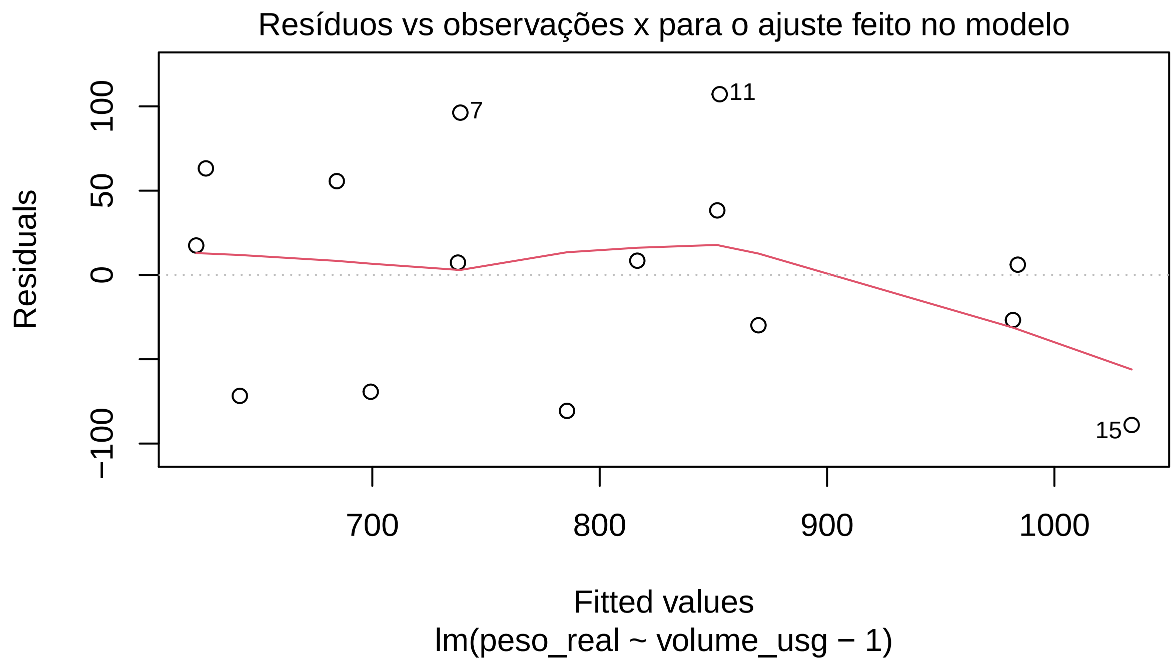
Novamente, os gráficos indicam que a observação 15 é um outlier. Refaremos o ajuste removendo a observação 15.

```r
dados4 <- dados1[-c(15), ]
ajuste <- ajustarModelo(dados4)
```

```
## [1] "O ajuste encontrou o coeficiente:"
## [1] "Beta: 1.06597728783179"
```




```
##
## Call:
## lm(formula = peso_real ~ volume_usg - 1, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -88.998 -49.559   7.344  46.963 107.218
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## volume_usg  1.06598    0.02156   49.44   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.11 on 14 degrees of freedom
## Multiple R-squared:  0.9943, Adjusted R-squared:  0.9939
## F-statistic:  2444 on 1 and 14 DF,  p-value: < 2.2e-16
```


Resíduos vs observações x para o ajuste feito no modelo
Residuals
100
50
0
−100
7
11
15
700
800
900
1000
Fitted values
lm(peso_real ~ volume_usg − 1)

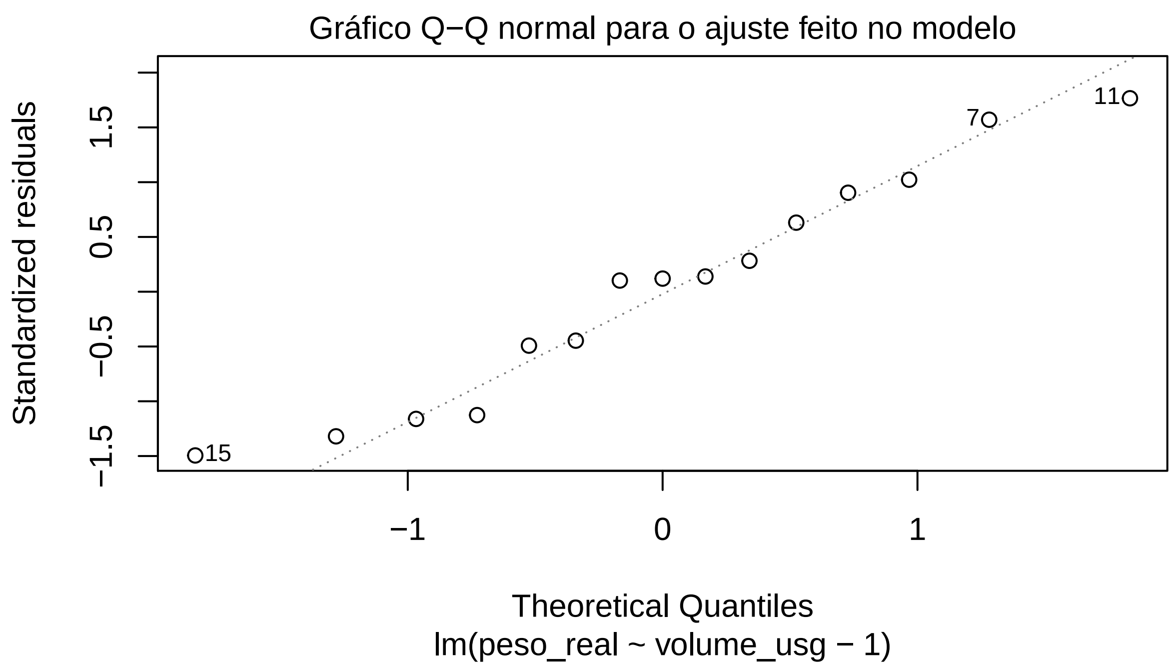
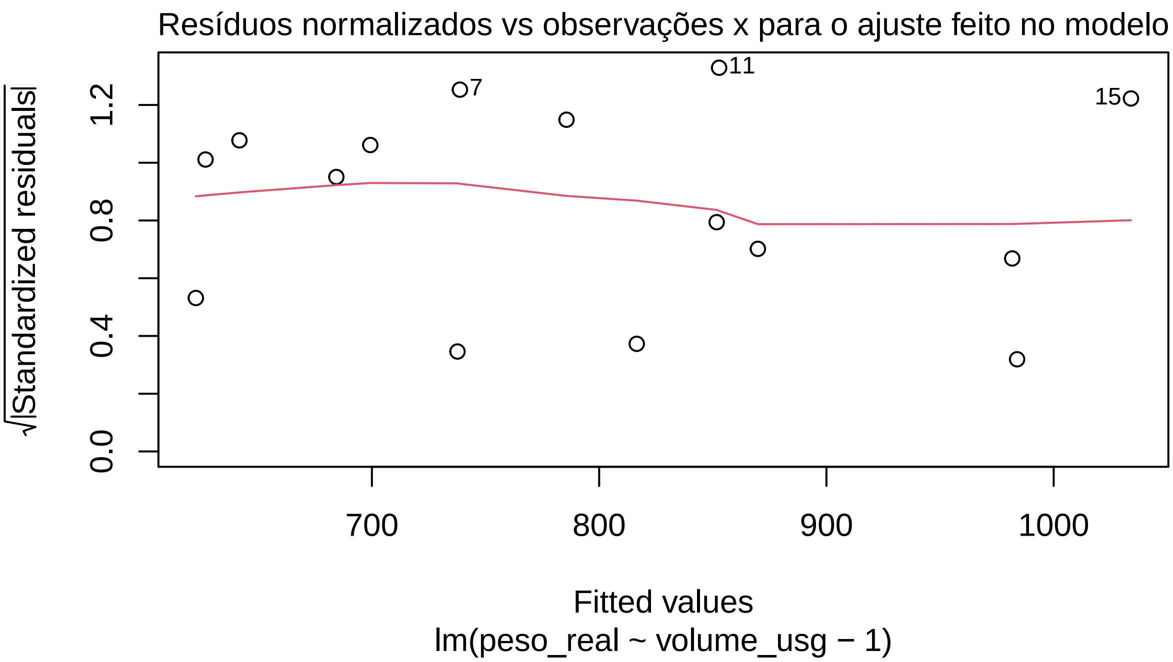
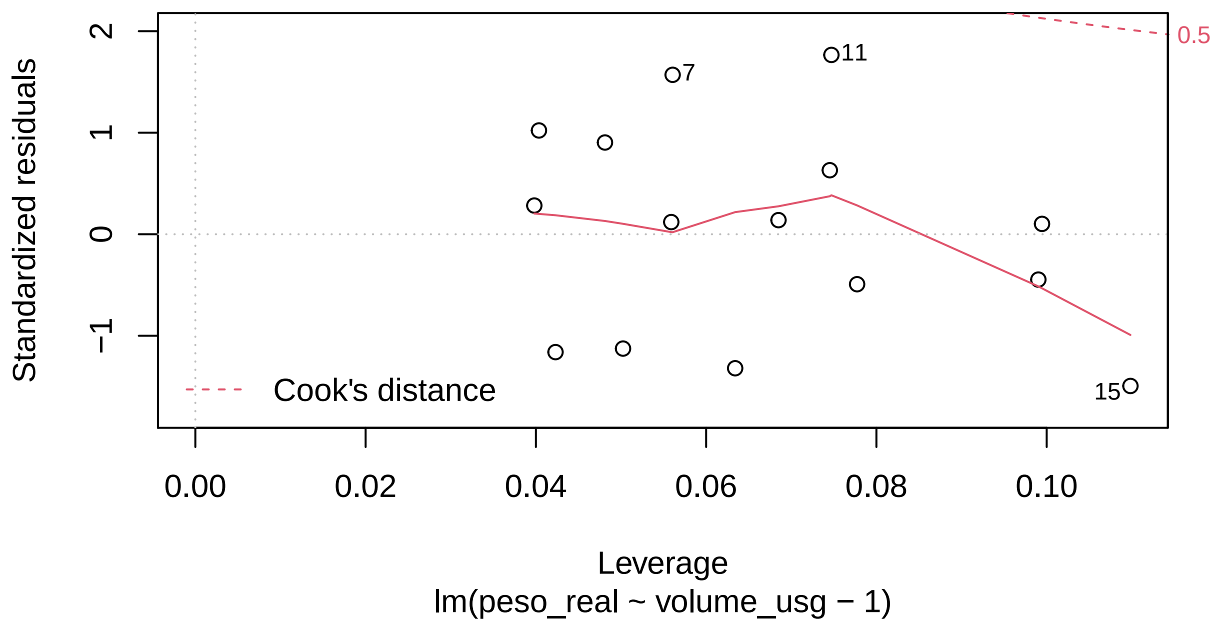


Gráfico Q−Q normal para o ajuste feito no modelo
Standardized residuals
1.5
0.5
−0.5
−1.5
11
7
15
−1
0
1
Theoretical Quantiles
lm(peso_real ~ volume_usg − 1)





As mesmas observações sobre a qualidade do modelo se aplicam. Os gráficos indicam que os resíduos possuem os valores dentro do esperado. Idealmente, o R^2 deveria estar próximo de 1, mas não está. Dessa forma, podemos concluir que o ajuste do modelo aproxima os dados, mas não estritamente. Assim, espera-se que o intervalo de confiança ao prever o peso real com base no volume seja grande.

vi)iv)

Construindo intervalos de confiança dos parâmetros:

```
confidence_intervals <- confint(ajuste)
rownames(confidence_intervals) <- c("Beta")
kable(confidence_intervals, caption="Intervalos de confiança para o ajuste dos parâmetros do modelo")
```

Tabela 3: Intervalos de confiança para o ajuste dos parâmetros do modelo

	2.5 %	97.5 %
Beta	1,02	1,11

vi)v)

A seguir, construiremos a tabela.

```
volumes <- c(600, 700, 800, 900, 1000)
df <- data.frame(volume_usg = volumes)
previsto <- predict(ajuste, df, interval='confidence')
previsto <- data.frame(previsto)
intervalo <- previsto$fit - previsto$lwr
previsto <- cbind(volume_usg = volumes, intervalo=previsto$fit, intervalo = intervalo)
colnames(previsto) <- c("Volume", "Peso previsto", "Intervalo de confiança de 95%")
kable(previsto, caption="Pesos previstos pelo modelo")
```

Tabela 4: Pesos previstos pelo modelo

Volume	Peso previsto	Intervalo de confiança de 95%
600	639,59	27,75
700	746,18	32,37
800	852,78	37,00
900	959,38	41,62
1.000	1.065,98	46,25

vi)vi)

Ambos os modelos satisfazem de forma similar as métricas mostradas na etapa (iii). Entretanto, observa-se na etapa (v) que o segundo modelo apresenta intervalos de confiança menores para suas predições de peso real. Dessa forma, o modelo sem intersecto demonstrou-se mais conveniente. Destacamos que o intervalo de confiança de 97,5% do parâmetro α no primeiro modelo era consideravelmente alto, o que poderia indicar que ele não possuia muita importância no modelo.

Exercício 2

i) O valor α, correspondente ao ponto onde a reta da regressão corta o eixo y quando $x = 0$, será nesse caso o valor médio dentre todas as médias das notas obtidas, tanto por alunos de escolas públicas quanto particulares. E β corresponde a metade da diferença entre a média das médias dos alunos na escola particular e pública.

ii) Sejam $\hat{\alpha}$ e $\hat{\beta}$ estimativa para α e β respectivamente, pelo Método dos Mínimos Quadrados temos:

```
medias <- read_xlsx("data/medias_escolas.xlsx")
x <- medias$x
y <- medias$y

x_bar <- mean(x)
y_bar <- mean(y)

beta_hat <- sum( (x-x_bar)*(y-y_bar) )/sum( (x-x_bar)^2 )
alpha_hat <- y_bar - beta_hat*x_bar
```

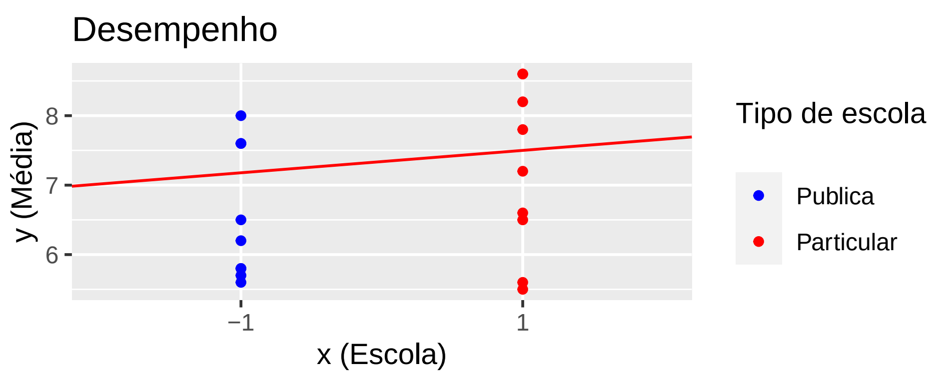
```
## [1] "Alpha estimado:  6.85555555555556"
```

```
## [1] "Beta estimado:  0.322222222222222"
```

Estimativa S^2 para σ^2:

```
n = length(x)
y_pred <- alpha_hat + beta_hat*x
residuos <- y-y_pred
SQRes <- sum(residuos^2)
S2 <- SQRes/(n-2)
```

```
## [1] "Variância estimada:  1.17222222222222"
```



iii) Avalie a qualidade do modelo através de técnicas de diagnóstico

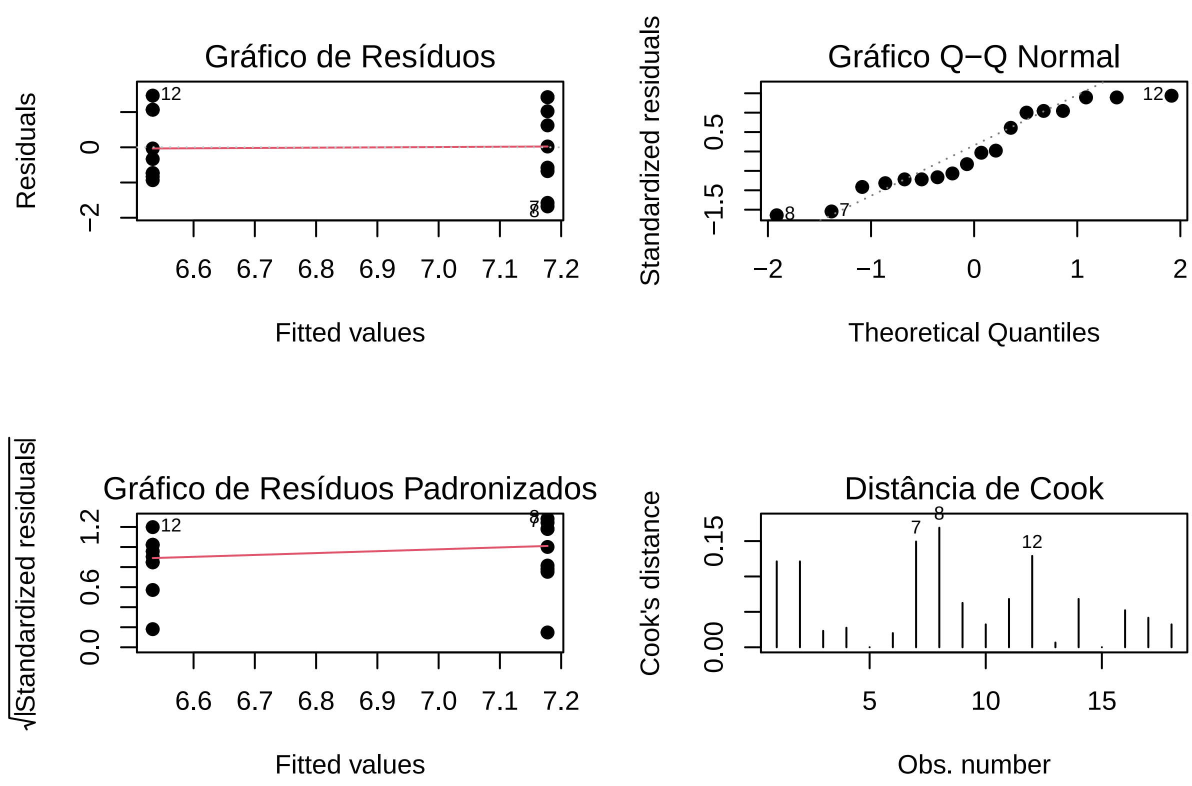
Através do coeficiênte de determinação:

```
y_pred <- alpha_hat + beta_hat*x
SQTot <- sum( (y-y_bar)^2 )
SQRes <- sum( (y-y_pred)^2 )

R2 <- 1-SQRes/SQTot
```

```
## [1] "Coeficiente R2:  0.0906152354272169"
```

Vemos que o modelo proposto explica apenas 9% da variância dos dados, formando um ajuste ruim.

```
modelo <- lm(medias, formula = y~x)
par(mfrow = c(2,2))
plot(modelo, which=c(1:4), pch=19, caption=plot_titles)
```



Analisando o Gráfico de Cook temos os pontos alavanca destacados (7, 8 e 12), indicando que são os pontos de maior influência na estimação dos parâmetros. No gráfico de Resíduos Padronizados vemos na parte inferior outliers, que não aparecem nos demais gráficos, também nota-se que a variância dos dados é aparentemente uniforme com a variação da variável explicativa, sugerindo homocedastidade. Já no Gráfico Q-Q Normal os pontos visualmente se aproximam de uma linha reta, indicando que a distribuição das médias das notas apresenta comportamento similar ao da distribuição normal.

iv) Supondo que e_i possui distribuição normal e não são correlacionados.

Usando que

$$Var(\hat{\alpha}) = \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

têm-se

```
var_alpha <- S2*sum(x^2)/(n*sum((x-x_bar)^2))
delta_alpha <- 1.96*sqrt(var_alpha/n)
delta_alpha
```

```
## [1] 0.1178931
```

Agora para $\hat{\beta}$, com:

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

obtemos:

```
var_beta <- S2/sum( (x-x_bar)^2 )
delta_beta <- 1.96*sqrt(var_beta/n)
delta_beta
```

```
## [1] 0.1178931
```

Disso, os limites para os intervalos de confiança de 95% são:

Tabela 5: Intervalo de confiança de 95% para α e β

Parâmetro	Valor esperado	limite inferior	limite superior
α	6.8556	6.7377	6.9734
β	0.3222	0.2043	0.4401

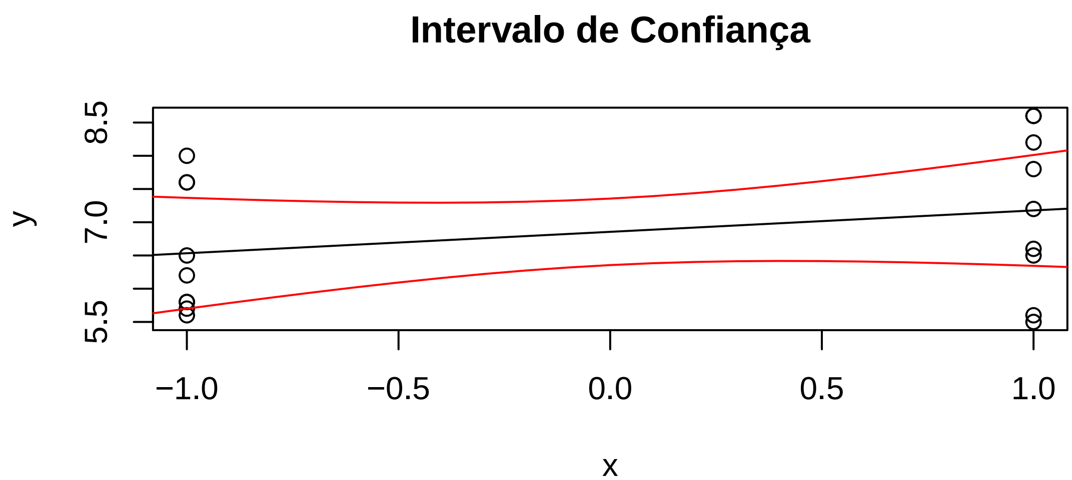
Obs.: os valores apresentados estão arredondados na quarta casa decimal.

v) Usando que o intervalo de 95% de confiança para a estimação das notas é:

$$\hat{y} \pm 1,96S\sqrt{\frac{1}{n} + \frac{(x_0 + \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

temos:

```r
desvio <- sqrt(S2)
x_val = (-11:11)/10 #pontos intermediários para plotar o gráfico
y_val = alpha_hat + beta_hat*x_val
delta <- 1.96*desvio*sqrt(1/n+(x_val-x_bar)^2/sum((x_val-x_bar)^2))

upper = y_val+delta
lower = y_val-delta

plot(x, y)
title("Intervalo de Confiança")
abline(a = alpha_hat, b = beta_hat)
lines(x_val, upper, col="red")
lines(x_val, lower, col="red")
```



Quando $x_0 \in \{-1, 1\}$ com $\bar{x} = 0$ temos:

```r
delta <- 1.96*desvio*sqrt(1/n+1/sum(x^2))
delta
```

```
## [1] 0.7073589
```

Tabela 6: Intervalos de confiança de 95% para o valor esperado das notas

Escola	$\hat{y}$	limite inferior	limite superior
Particular	7.178	6.4704	7.8851
Pública	6.534	5.8260	7.2407

Obs.: os valores apresentados estão arredondados na quarta casa decimal.

vi)

vii)

a) Nesse caso α é o valor de y em $x = 0$ da reta da regressão, ou seja, correspondente a média das médias dos alunos de escola pública. Já β é a diferença entre a média das médias dos alunos de escola particular e pública.

b) Sendo $\hat{\alpha}$ e $\hat{\beta}$ estimativas de α e β pelo Método de Mínimos Quadrados:

```
medias <- read_xlsx("data/medias_escolas.xlsx")
medias[medias == -1] <- 0
x <- medias$x
y <- medias$y

x_bar <- mean(x)
y_bar <- mean(y)

beta_hat <- sum((x-x_bar)*(y-y_bar))/sum((x-x_bar)^2)
alpha_hat <- y_bar-beta_hat*x_bar
```

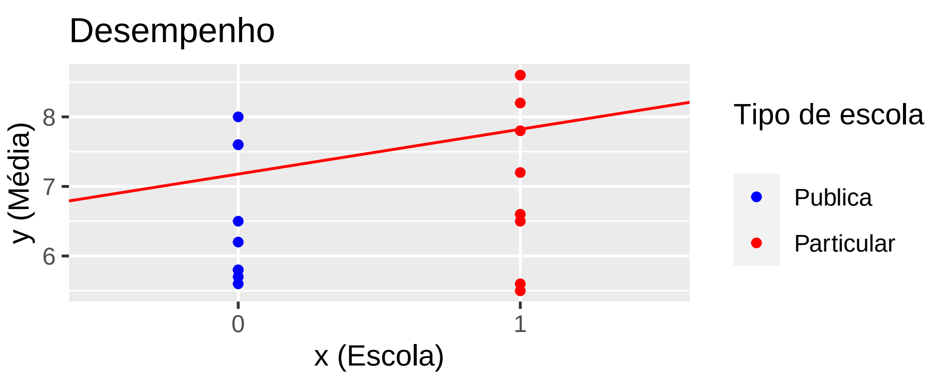
```
## [1] "Alpha estimado:  6.53333333333333"
```

```
## [1] "Beta estimado: 0.644444444444444"
```

Estimativa S^2 para σ^2:

```
n = length(x)
y_pred <- alpha_hat+beta_hat*x
residuos <- y-y_pred
SQRes <- sum(residuos^2)
S2 <- 1/(n-2)*SQRes
```

```
## [1] "Variância estimada:  1.17222222222222"
```



c) Qualidade do ajuste

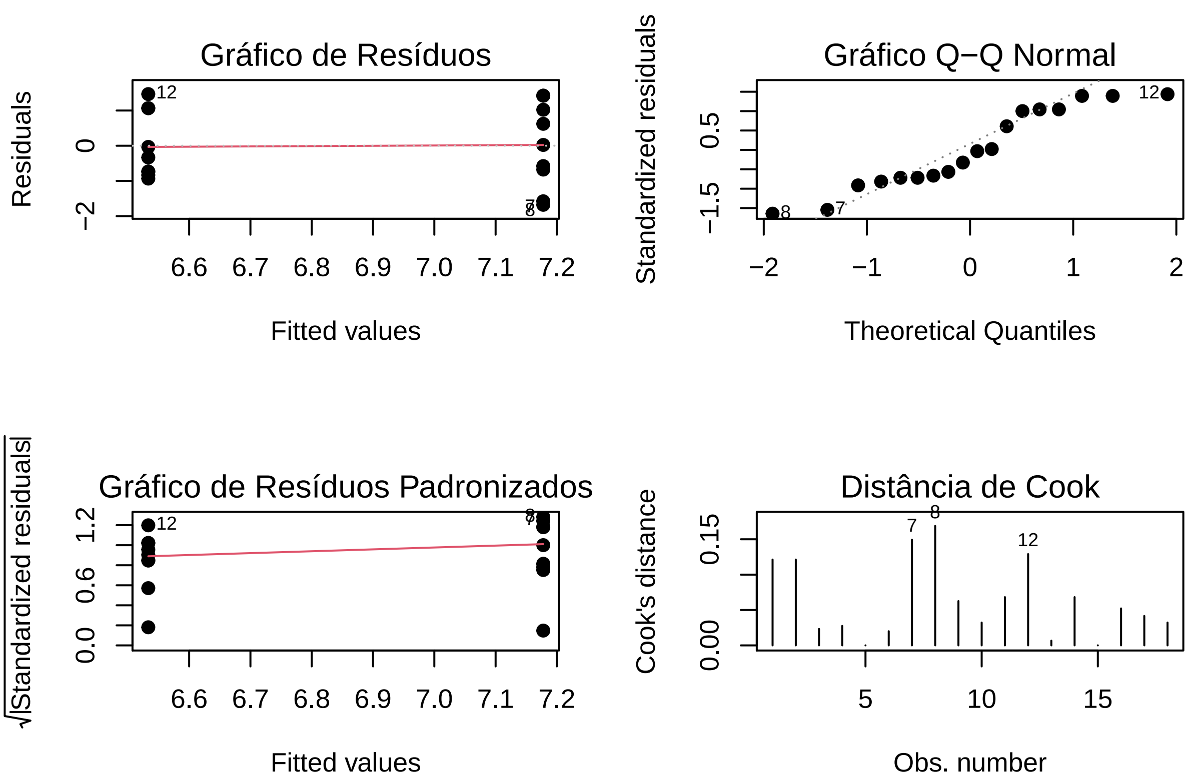
Através do coeficiênte de determinação:

```r
y_pred <- alpha_hat+beta_hat*x
SQTot <- sum((y-y_bar)^2)
SQRes <- sum((y-y_pred)^2)

R2 <- 1-SQRes/SQTot
```

```
## [1] "Coeficiente R2:  0.0906152354272169"
```

Obtemos o mesmo coeficiente de determinação que no item (iii), tendo, como antes, pouca explicação dos dados pelo modelo.

```r
modelo <- lm(medias, formula = y~x)
par(mfrow = c(2,2))
plot(modelo, which=c(1:4), pch=19, caption=plot_titles)
```



Assim como no item (iii), temos no Gráfico de Cook os mesmos pontos alavanca (7, 8 e 12), influenciando notavelmente na estimação dos parâmetros. No gráfico de Resíduos Padronizados vemos na parte inferior outliers em torno do valor 0.2, que não aparecem nos demais gráficos, ainda no Gráfico de Resíduos Padronizados notamos a homocedastidade. E o Gráfico Q-Q Normal sugere que a distribuição das médias das notas apresenta comportamento similar ao da distribuição normal.

d) Analogamente ao item (iv), supondo que e_i possui distribuição normal e não são correlacionados. Temos:

```
var_alpha <- S2*sum(x^2)/(n*sum((x-x_bar)^2))
delta_alpha <- 1.96*sqrt(var_alpha/n)
delta_alpha
```

```
## [1] 0.1667261
```

```
var_beta <- S2/sum( (x-x_bar)^2 )
delta_beta <- 1.96*sqrt(var_beta/n)
delta_beta
```

```
## [1] 0.2357863
```

Portanto, os limites para os intervalos de confiança de 95% são:

Tabela 7: Intervalo de confiança de 95% para α e β

Parâmetro	Valor esperado	limite inferior	limite superior
α	6.5334	6.3667	6.7001
β	0.6444	0.4087	0.8802

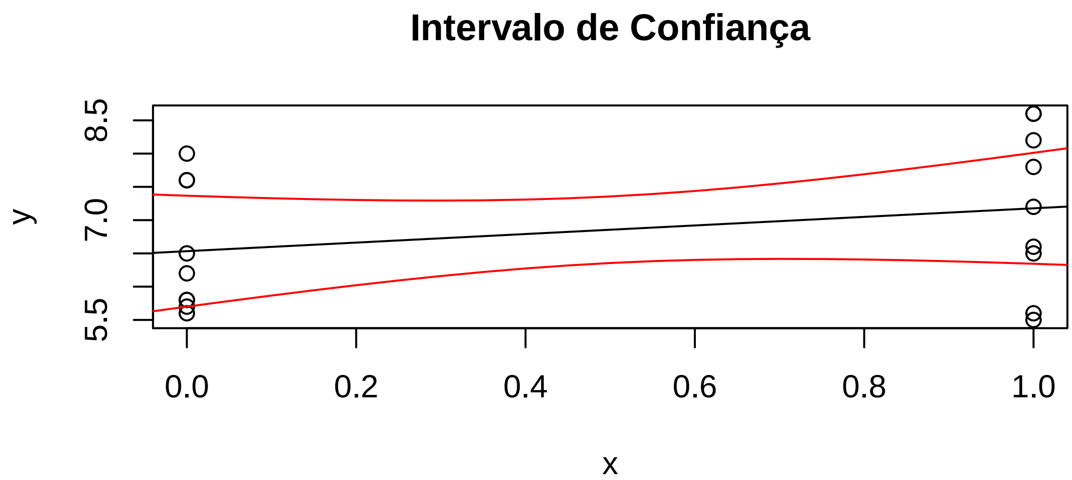
Obs.: os valores apresentados estão arredondados na quarta casa decimal.

e) Usando que o intervalo de 95% de confiança para a estimação das notas é:

$$\hat{y} \pm 1,96S\sqrt{\frac{1}{n} + \frac{(x_0 + \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

temos:

```
desvio <- sqrt(S2)
x_val = (-1:21)/20 # pontos intermediários para plotar o gráfico
y_val = alpha_hat + beta_hat*x_val
delta <- 1.96*desvio*sqrt(1/n+(x_val-x_bar)^2/sum((x_val-x_bar)^2))

upper = y_val+delta
lower = y_val-delta

plot(x, y)
title("Intervalo de Confiança")
abline(a = alpha_hat, b = beta_hat)
lines(x_val, upper, col="red")
lines(x_val, lower, col="red")
```



Quando $x_0 \in \{-1, 1\}$ com $\bar{x} = 0$ temos:

```
delta <- 1.96*desvio*sqrt(1/n+1/sum(x^2))
delta
```

```
## [1] 0.8663341
```

Tabela 8: Intervalos de confiança de 95% para o valor esperado das notas

Escola	$\hat{y}$	limite inferior	limite superior
Particular	7.178	6.3114	8.0441
Pública	6.534	5.6667	7.3997

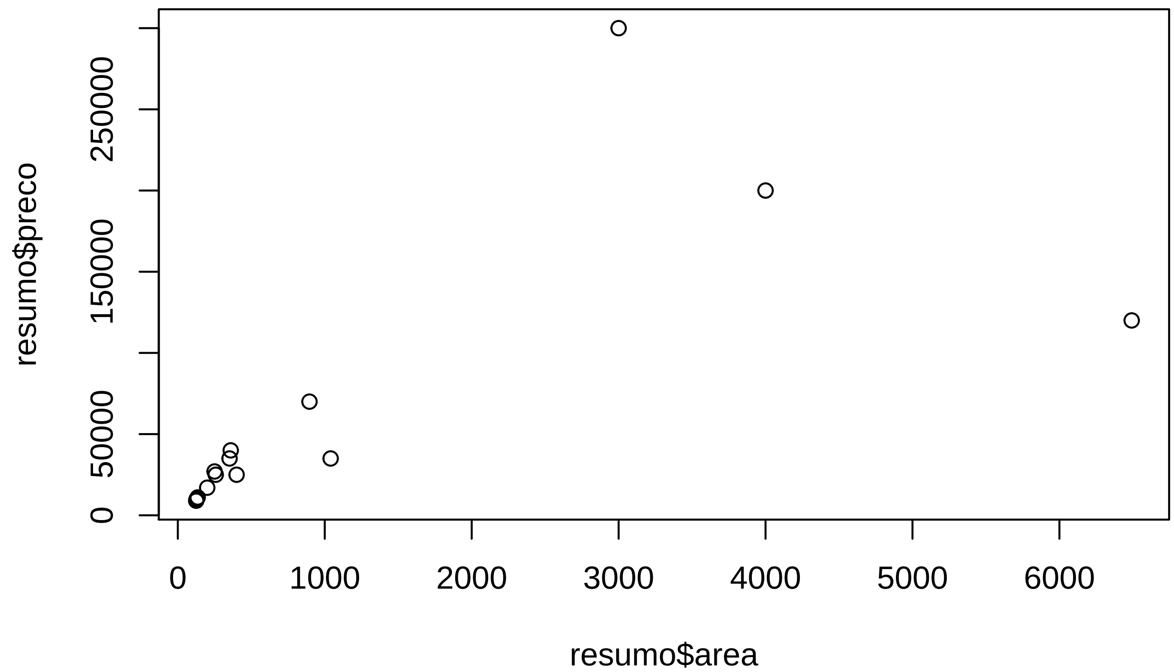
Obs.: os valores apresentados estão arredondados na quarta casa decimal.

Exercício 3

i)

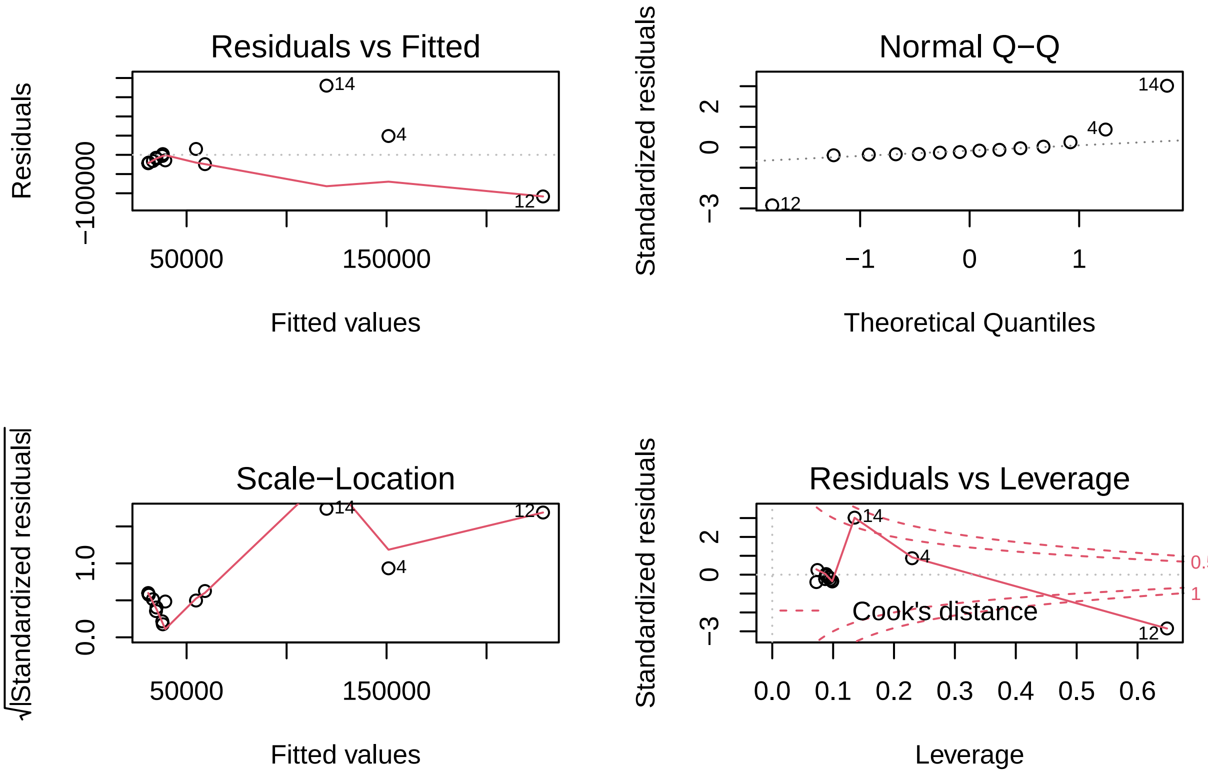
```
imovel <- c("1",  "2",  "3",  "4" , "5",  "6"  ,"7" , "8" , "9" , "10", "11", "12", "13", "14")

area <- c(128 , 125,  200, 4000,  258,  360,  896,  400,  352,  250,  135, 6492, 1040, 3000)

preco <- c(10000,  9000,  17000, 200000,  25000,  40000,  70000,  25000,  35000,  27000,  11000, 120000
35000, 300000)

resumo <- cbind(imovel, area, preco)
resumo <- as.data.frame(apply(resumo, 2, as.numeric))
```

```
plot(resumo$area,resumo$preco)
```



ii)

```
mod <- lm(preco ~ area, data=resumo)

par(mfrow=c(2,2))
plot(mod)
```



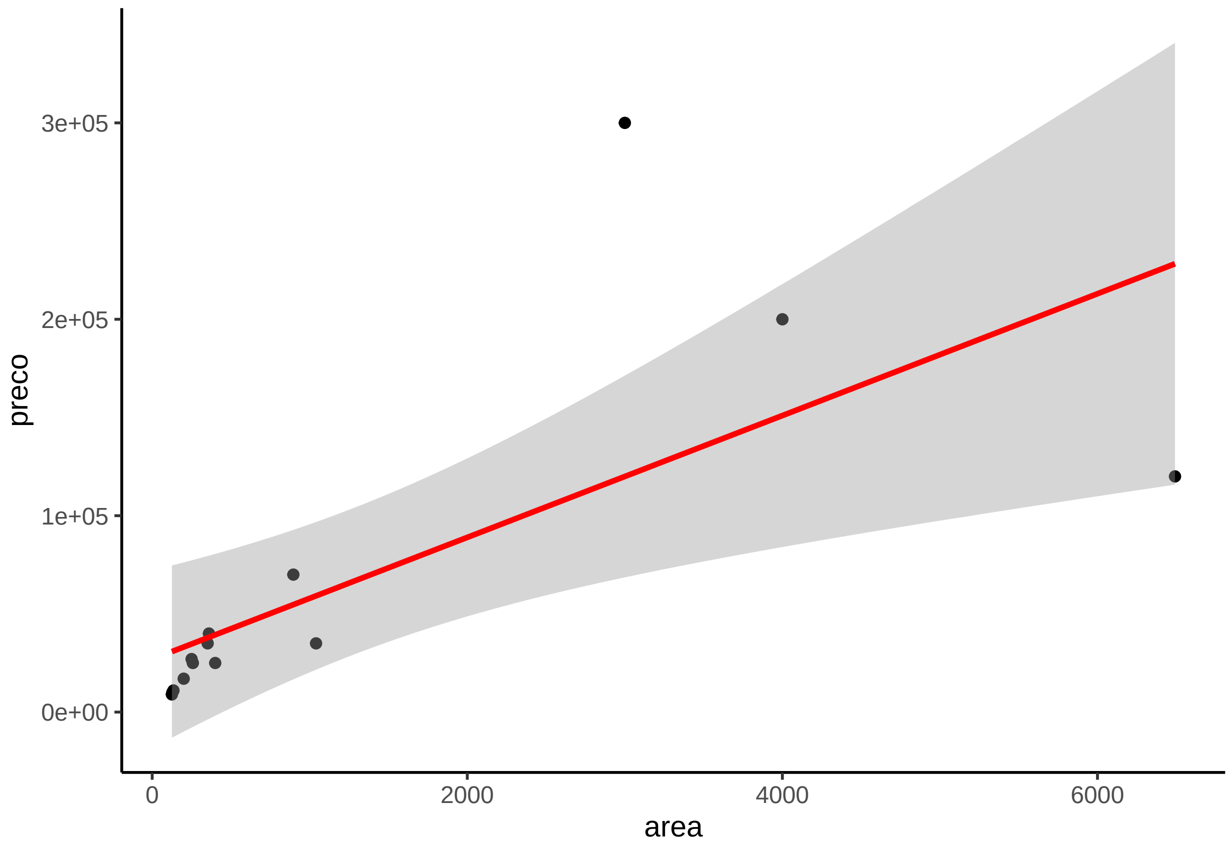

```r
summary(mod)
```

```
##
## Call:
## lm(formula = preco ~ area, data = resumo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -108260  -20708  -12137     713  180031
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26934.568  20758.379   1.298  0.21884
## area           31.011      9.306   3.332  0.00597 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64100 on 12 degrees of freedom
## Multiple R-squared:  0.4806, Adjusted R-squared:  0.4374
## F-statistic: 11.11 on 1 and 12 DF,  p-value: 0.005971
```

```r
ggplot(data = resumo, mapping = aes(x = area, y=preco))+
  geom_point()+
```

```
  geom_smooth(method = "lm", col = "red")+
  theme_classic()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



$$y = 31.011 * x + 26935$$

iii)

$$y = 52051 \log(x) - 259408$$

$$R^2 = 0.642$$

iv)

```r
predict(mod, data.frame(area = 200))
```

```
##        1
## 33136.84
```

```r
predict(mod, data.frame(area = 500))
```

```
##        1
## 42440.24
```

```r
predict(mod, data.frame(area = 1000))
```

```
##        1
## 57945.91
```

Exercício 4

Exercício 15

a)

$SSE(\hat{\beta}) = \sum_{i=1}^{n} \hat{e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}x_i)^2$

$\frac{dSSE(\hat{\beta})}{d\hat{\beta}} = \sum_{i=1}^{n} 2(y_i - \hat{\beta}x_i)(-x_i) = -2\sum_{i=1}^{n}(y_i x_i - \hat{\beta}x_i^2)$

$\frac{dSSE(\hat{\beta})}{d\hat{\beta}} = -2\sum_{i=1}^{n}(y_i x_i) + 2\hat{\beta}\sum_{i=1}^{n} x_i^2 = 0, \text{para minimizar}$

$2\sum_{i=1}^{n}(y_i x_i) = 2\hat{\beta}\sum_{i=1}^{n} x_i^2$

$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i x_i)}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} SSE(\beta)$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\beta}x_i^2)$

b)

Para o modelo de regressão linear simples classico, onde
$\hat{e}_i \sim Normal(0, \sigma^2)$ temos $\hat{\beta} \sim t - student(n-1)$

c)

Sendo t para $\gamma/2$ com n-1 graus de liberdade, temos:
$IC(\beta) = \hat{\beta} \pm t\sqrt{S^2} = \beta^2 \pm t\frac{\sum_{i=1}^{n}(y_i - \hat{\beta}x_i)}{\sqrt{n-1}}$

Exercício 16