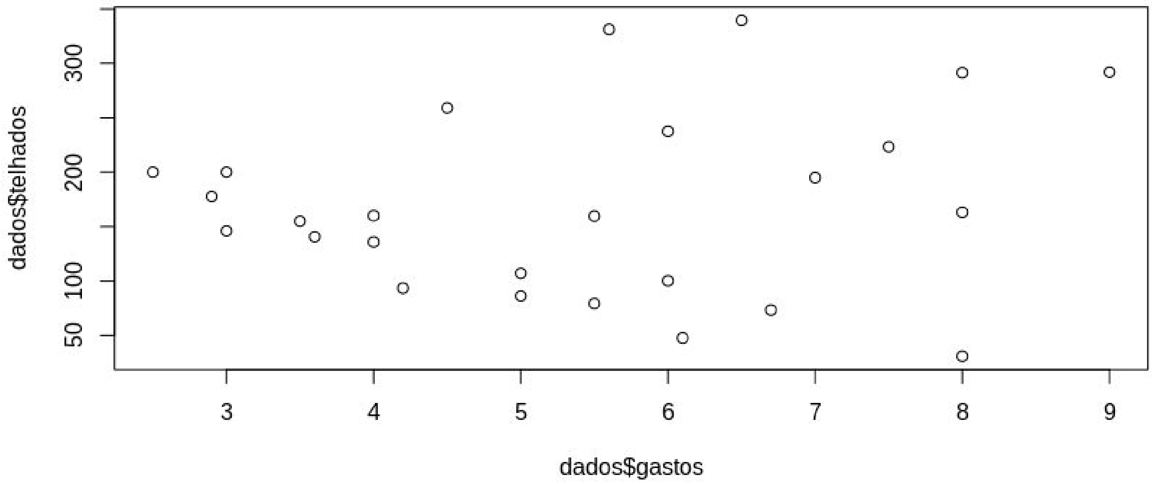
Laboratório 09 - Estatística Aplicada - Manhã
Aluno: Marcus Vinícius Leite Costa
Matrícula: 116110728

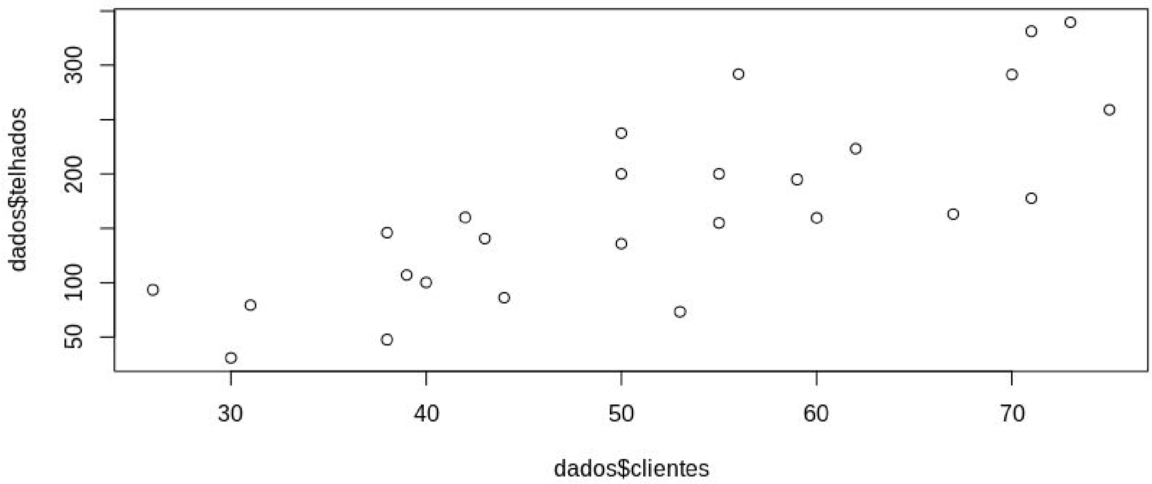
1)

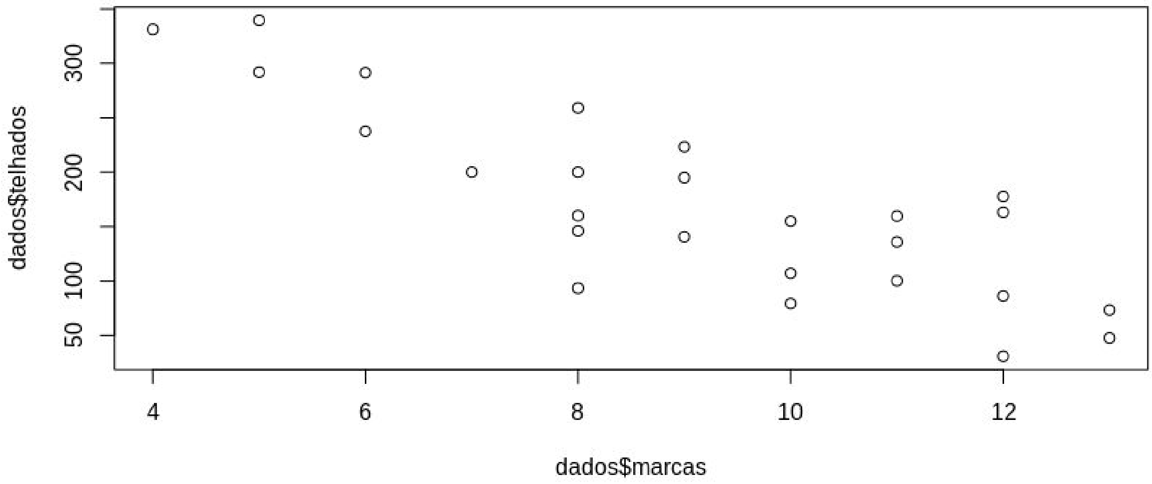
a) Na letra A, foi-se utilizado os seguintes comandos para gerar os diagramas de dispersão entre as variáveis:

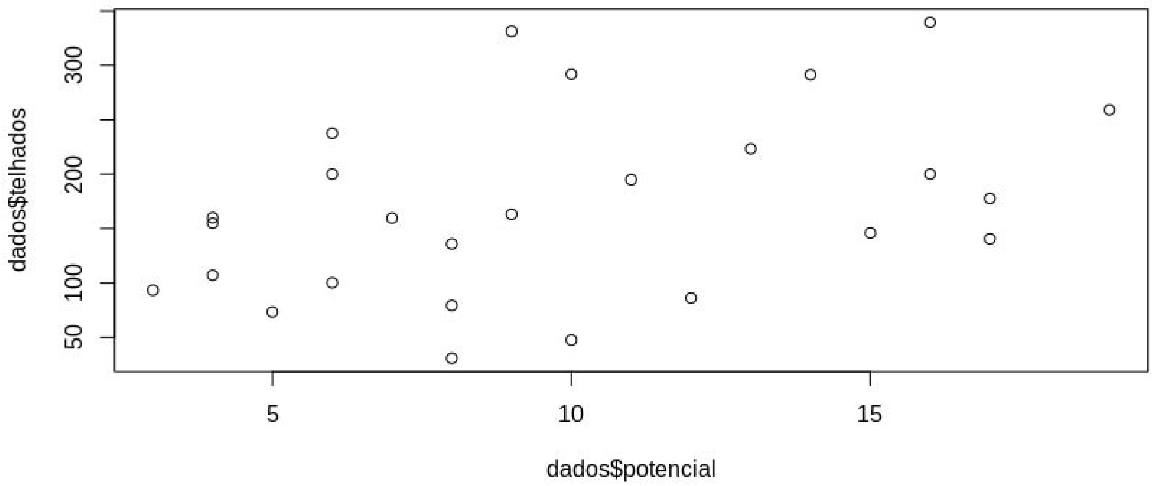
```
# QUESTAO 01

dados <- scan("vendas.dat",what=list(telhados=0, gastos=0,
clientes=0, marcas=0, potencial=0))
attach(dados)

# Letra A

plot(x = dados$gastos , y = dados$telhados)

plot(x = dados$clientes , y = dados$telhados)

plot(x = dados$marcas , y = dados$telhados)

plot(x = dados$potencial , y = dados$telhados)
```

Os diagramas de dispersão gerados foram os seguintes:


dados$telhados
50
100
200
300
3
4
5
6
7
8
9
dados$gastos


dados$telhados
50
100
200
300
30
40
50
60
70
dados$clientes





b) Na letra B, foi-se utilizado um comando que plota uma matriz de correlação entre todas as variáveis, o comando foi o seguinte:

```
cor(matrix(unlist(dados), nrow=26))
```

E a matriz obtida foi:

```
           [,1]        [,2]       [,3]        [,4]       [,5]
[1,]  1.0000000  0.15886840  0.7828441 -0.83298059  0.4072695
[2,]  0.1588684  1.00000000  0.1725506 -0.03825157 -0.0705647
[3,]  0.7828441  0.17255062  1.0000000 -0.32428751  0.4682129
[4,] -0.8329806 -0.03825157 -0.3242875  1.00000000 -0.2020604
[5,]  0.4072695 -0.07056470  0.4682129 -0.20206035  1.0000000
```

Onde a posição 1,2,3,4,5 é referente às variáveis telhados, gastos, clientes, marcas e potencial respectivamente.

c) Na letra C, primeiramente foi-se utilizado os comandos para adicionar as variáveis ao modelo em ordem de correlação (calculada anteriormente)

```
lm1 <- lm(telhados ~ marcas)
lm2 <- lm(telhados ~ marcas+clientes)
lm3 <- lm(telhados ~ marcas+clientes+potencial)
lm4 <- lm(telhados ~ gastos+clientes+marcas+potencial)
```

Após obter-se os modelos, foi-se utilizado o comando summary para gerar as tabelas anova, os resultados obtidos foram os seguintes:

```
> summary(lm1)

Call:
lm(formula = telhados ~ marcas)

Residuals:
     Min        1Q    Median      3Q     Max
-107.133  -31.444     4.195   32.570   86.178

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept)  418.854   34.990   11.971 1.31e-11 ***
marcas      -27.278    3.699   -7.375 1.29e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.75 on 24 degrees of freedom
Multiple R-squared:  0.6939,    Adjusted R-squared:  0.6811
F-statistic: 54.39 on 1 and 24 DF,  p-value: 1.29e-07

> summary(lm2)

Call:
lm(formula = telhados ~ marcas + clientes)

Residuals:
     Min        1Q    Median      3Q     Max
```

```
-18.4136  -6.1499  -0.5683   6.2472  20.3185

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 186.6940  12.2587   15.23 1.66e-13 ***
marcas     -21.1930   0.8028  -26.40  < 2e-16 ***
clientes   3.4081     0.1458   23.37  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.803 on 23 degrees of freedom
Multiple R-squared:  0.9876,    Adjusted R-squared:  0.9866
F-statistic: 918.3 on 2 and 23 DF,  p-value: < 2.2e-16

> summary(lm3)

Call:
lm(formula = telhados ~ marcas + clientes + potencial)

Residuals:
     Min        1Q   Median      3Q    Max
-17.5324  -5.5272  -0.5563   5.7107  20.4942

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 186.1147  12.5759  14.799 6.44e-13 ***
marcas     -21.1735   0.8194 -25.840  < 2e-16 ***
clientes   3.3798     0.1650  20.488 8.05e-16 ***
potencial  0.1892     0.4791   0.395   0.697
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.988 on 22 degrees of freedom
Multiple R-squared:  0.9877,    Adjusted R-squared:  0.986
F-statistic: 589.8 on 3 and 22 DF,  p-value: < 2.2e-16

> summary(lm4)

Call:
lm(formula = telhados ~ gastos + clientes + marcas + potencial)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-19.0906  -5.9796   0.8968   6.5667  14.7985 

Coefficients:
           Estimate Std. Error t value Pr(>|t|)    
(Intercept) 178.3203   12.9603  13.759 5.62e-12 ***
gastos      1.8071     1.0810   1.672   0.109    
clientes    3.3178     0.1629  20.368 2.60e-15 ***
marcas      -21.1850   0.7879 -26.887  < 2e-16 ***
potencial   0.3245     0.4678   0.694   0.495    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.604 on 21 degrees of freedom
Multiple R-squared:  0.9892,	Adjusted R-squared:  0.9871 
F-statistic: 479.1 on 4 and 21 DF,  p-value: < 2.2e-16
```

d) O melhor modelo obtido foi aquele após acrescentar as variáveis marcas e clientes por possuir o menor p-valor, com uma baixa quantidade de resíduos.

2)

a) Os vetores X e Y foram obtidos a partir dos seguintes comandos:

```
# QUESTAO 2

# Letra A



data2 <- matrix(c(37.310,
37.380,34.135,36.985,38.715,40.620,39.200,40.320,10,5,3,6,8,20,8,1
4,2,6,1,5,8,0,4,6,16,16,12,14,16,12,18,17), nrow = 8 , ncol =4 ,
byrow = FALSE)


vetor_x = matrix(c(1,1,1,1,1,1,1,1,37.310,
37.380,34.135,36.985,38.715,40.620,39.200,40.320,10,5,3,6,8,20,8,1
4,2,6,1,5,8,0,4,6,16,16,12,14,16,12,18,17), nrow = 8 , ncol =5 ,
```

```
byrow = FALSE)

vetor_y = c(37.310,
37.380,34.135,36.985,38.715,40.620,39.200,40.320)
```

Os resultados dos vetores X e Y, respectivamente, foram os seguintes:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	37.310	10	2	16
[2,]	1	37.380	5	6	16
[3,]	1	34.135	3	1	12
[4,]	1	36.985	6	5	14
[5,]	1	38.715	8	8	16
[6,]	1	40.620	20	0	12
[7,]	1	39.200	8	4	18
[8,]	1	40.320	14	6	17

[1] 37.310 37.380 34.135 36.985 38.715 40.620 39.200 40.320

b) A equação para calcular a reta foi obtida a partir da função:

```
beta <- solve(t(vetor_x) %*% vetor_x) %*% (t(vetor_x) %*% vetor_y)

modelo_matriz = function(beta1,x1,x2,x3){beta[1,] + beta[2,]*x1 +
beta[3,]*x2 + beta[4,]*x3}
```

c) O salário previsto, utilizando a seguinte função, foi 36.20141

modelo_matriz(beta[1,],7,3,12)

```
modelo_matriz = function(beta1,x1,x2,x3){beta[1,] + beta[2,]*x1 +
beta[3,]*x2 + beta[4,]*x3}
```

d) Chamando a função modelo_matriz, foi-se obtido o y_chapeu da seguinte forma:

```
y_chapeu = c(modelo_matriz(beta =
beta[1,],vetor_x[1,2],vetor_x[1,3],vetor_x[1,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[2,2],vetor_x[2,3],vetor_x[2,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[3,2],vetor_x[3,3],vetor_x[3,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[4,2],vetor_x[4,3],vetor_x[4,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[5,2],vetor_x[5,3],vetor_x[5,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[6,2],vetor_x[6,3],vetor_x[6,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[7,2],vetor_x[7,3],vetor_x[7,4]),
           modelo_matriz(beta =
beta[1,],vetor_x[7,2],vetor_x[8,3],vetor_x[8,4]))
```

E os resultados para o vetor y e os dados do modelo são os seguintes:

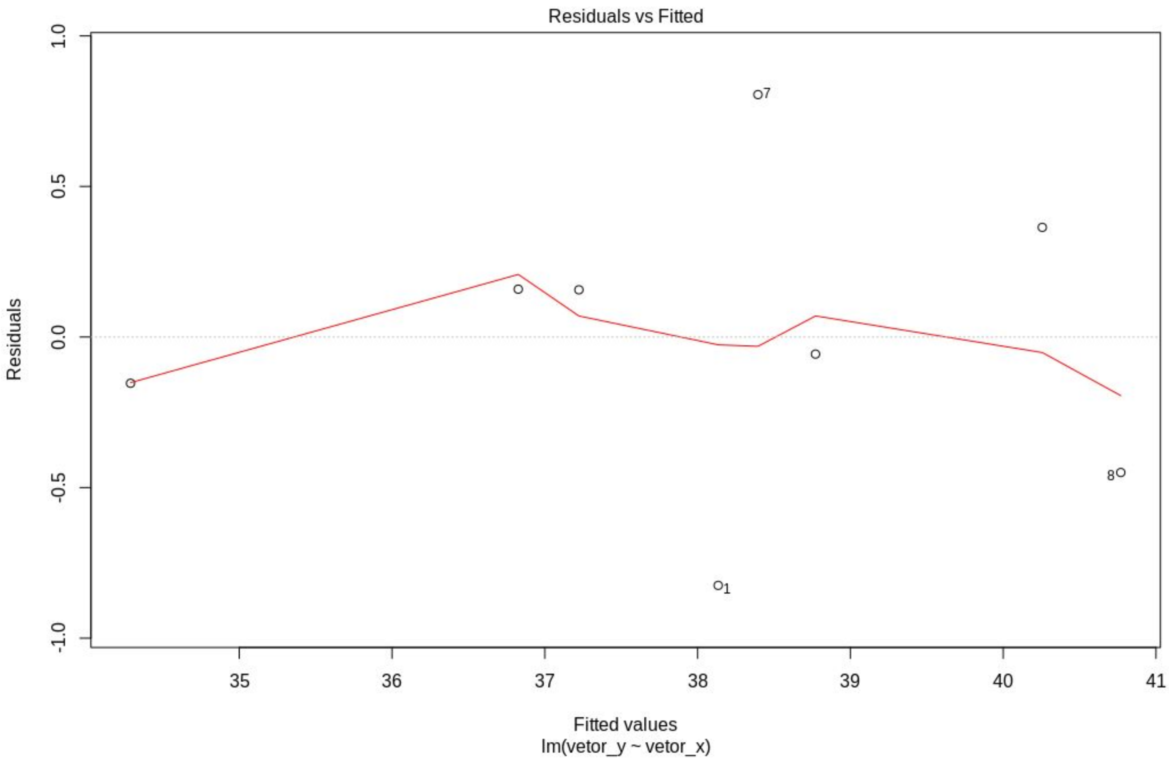
```
> vetor_y

[1] 37.310 37.380 34.135 36.985 38.715 40.620 39.200 40.320
> y_chapeu

[1] 38.13476 37.22318 34.28852 36.82610 38.77165 40.25591 38.39505 38.58335
```

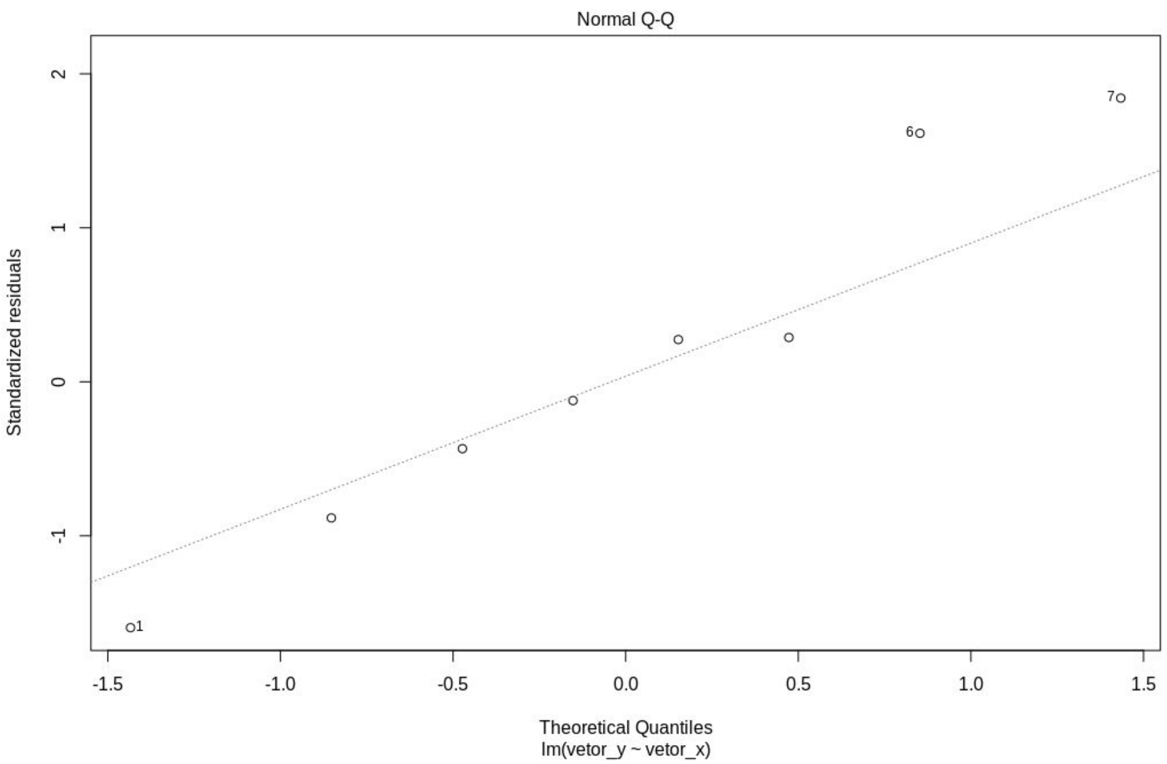
É possível observar que, os dados preditos estão bem próximos dos originais, logo, podemos deduzir que o modelo se ajusta bem aos dados.

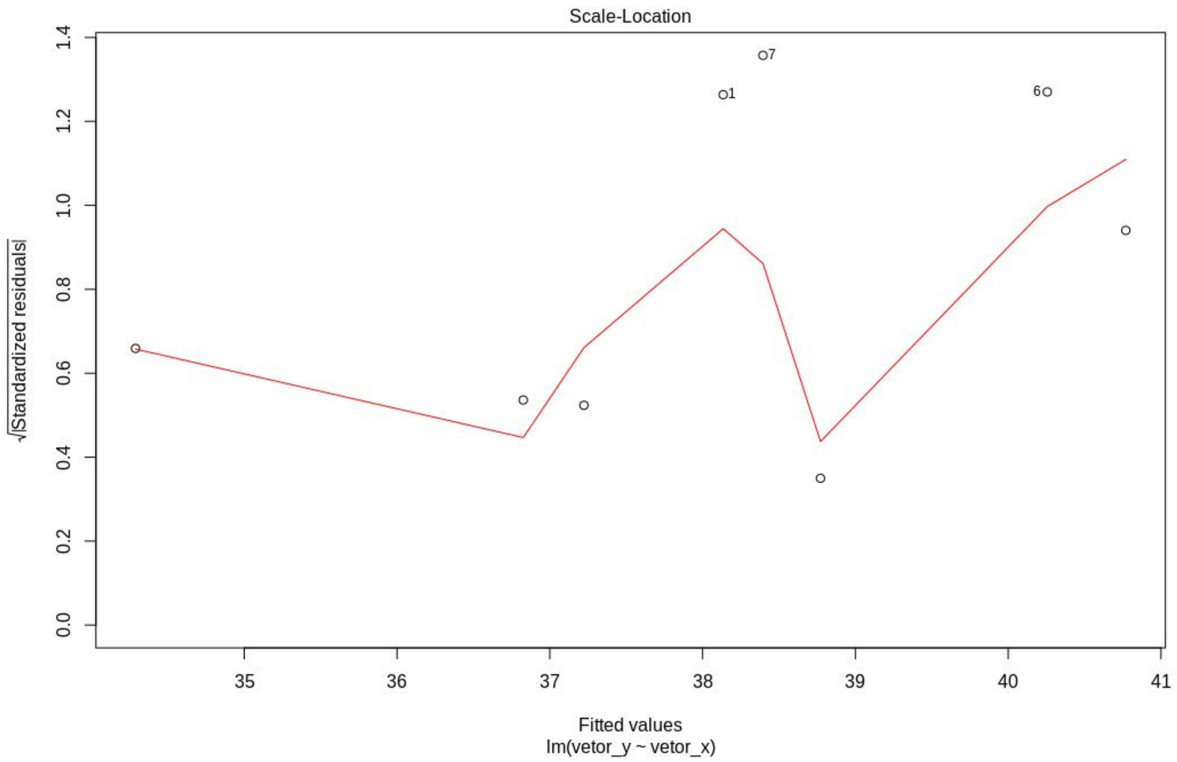
e) Os resíduos obtidos foram os seguintes:

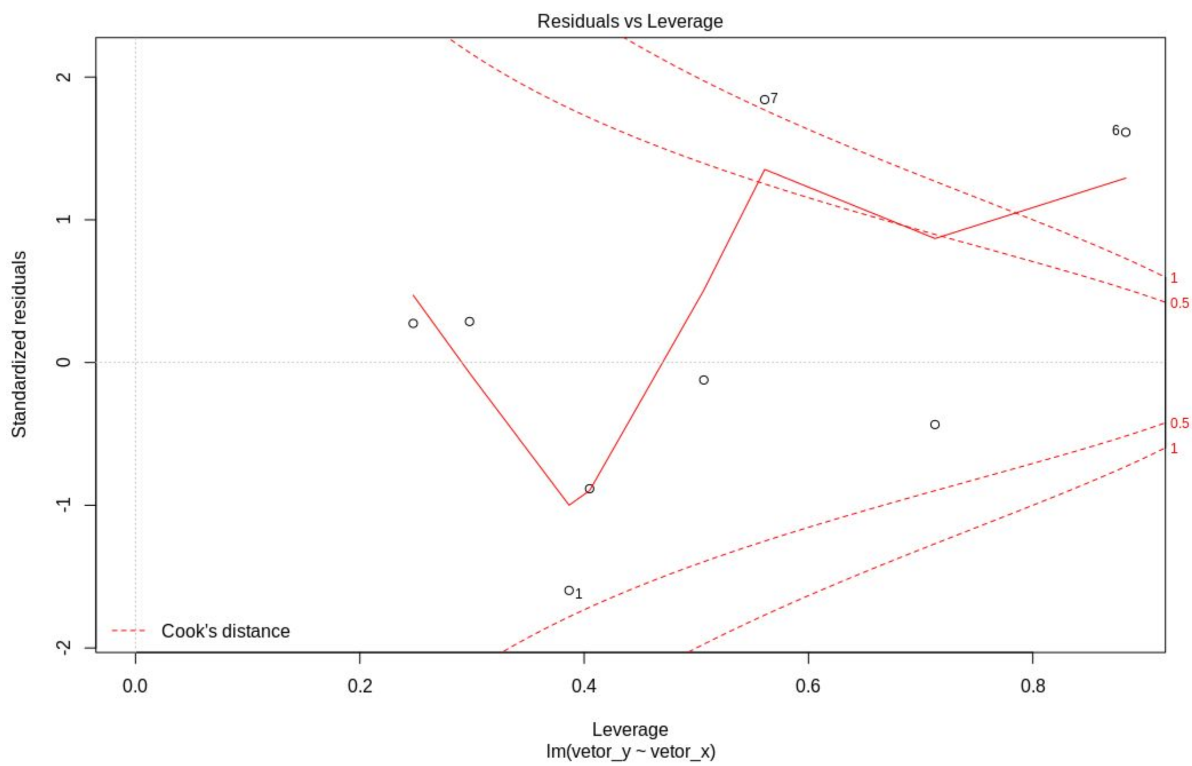
[1] -0.82476456 0.15682034 -0.15352139 0.15889720 -0.05665337 0.36409294 0.80495179 1.73664921

E ao rodar o Teste de Kolmogorov-Smirnov no modelo, foi-se observado os seguintes gráficos:


Residuals vs Fitted
Residuals
Fitted values
lm(vetor_y ~ vetor_x)
1.0
0.5
0.0
-0.5
-1.0
35
36
37
38
39
40
41
7
8
1


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(vetor_y ~ vetor_x)
2
1
0
-1
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
7
6
1





Ao observar os resultados dos resíduos, e também, os gráficos (pressupostos teóricos), podemos deduzir que o modelo se adaptou bem aos dados. Essa dedução é feita observando, primeiramente, o valor dos resíduos, que normalmente são baixos, e também, observando-se os gráficos, como por exemplo, no Q-Q Plot, que mostra claramente que não existem pontos distantes da curva.