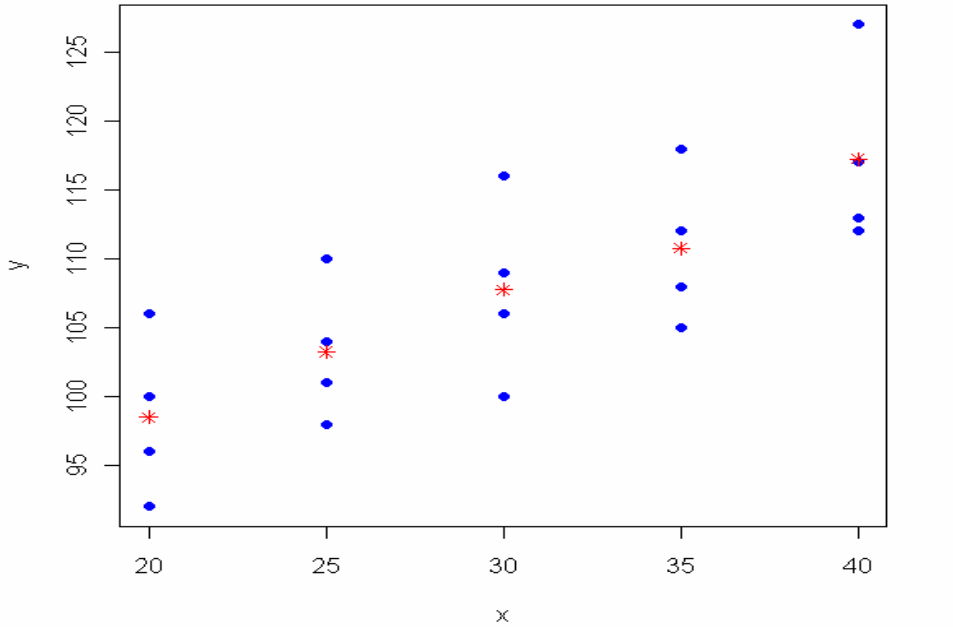
```
################################################################################
 CE071: ANALISE DE REGRESSAO LINEAR                    PROFa: SUELY RUIZ GIOLO

                              CODIGOS R VERSAO 2.10
                        EXEMPLOS: REGRESSÃO LINEAR SIMPLES
################################################################################

=======================================================================
EXEMPLO 1    Fonte: Bussab (1988).
Y = tempo de reação a certo estímulo (em segundos)
X = idade (em anos)
=======================================================================

ex1<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo1s.txt",h=T)
attach(ex1)
ex1

   tempo idade
1     96    20
2     92    20
3    106    20
4    100    20
5     98    25
6    104    25
7    110    25
8    101    25
9    116    30
10   106    30
11   109    30
12   100    30
13   112    35
14   105    35
15   118    35
16   108    35
17   113    40
18   112    40
19   127    40
20   117    40
```

```
x<-idade
y<-tempo
plot(x,y, pch=16, col=4)
points(c(20,25,30,35,40),c(98.5,103.25,107.75,110.75,117.25), pch=8, col=2)
```




```
cor(x,y)
[1] 0.7680814
cor.test(x,y)

        Pearson's product-moment correlation

data:  x and y
t = 5.0889, df = 18, p-value = 7.662e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
```

```
 0.4931929 0.9035073
sample estimates:
      cor
0.7680814
```

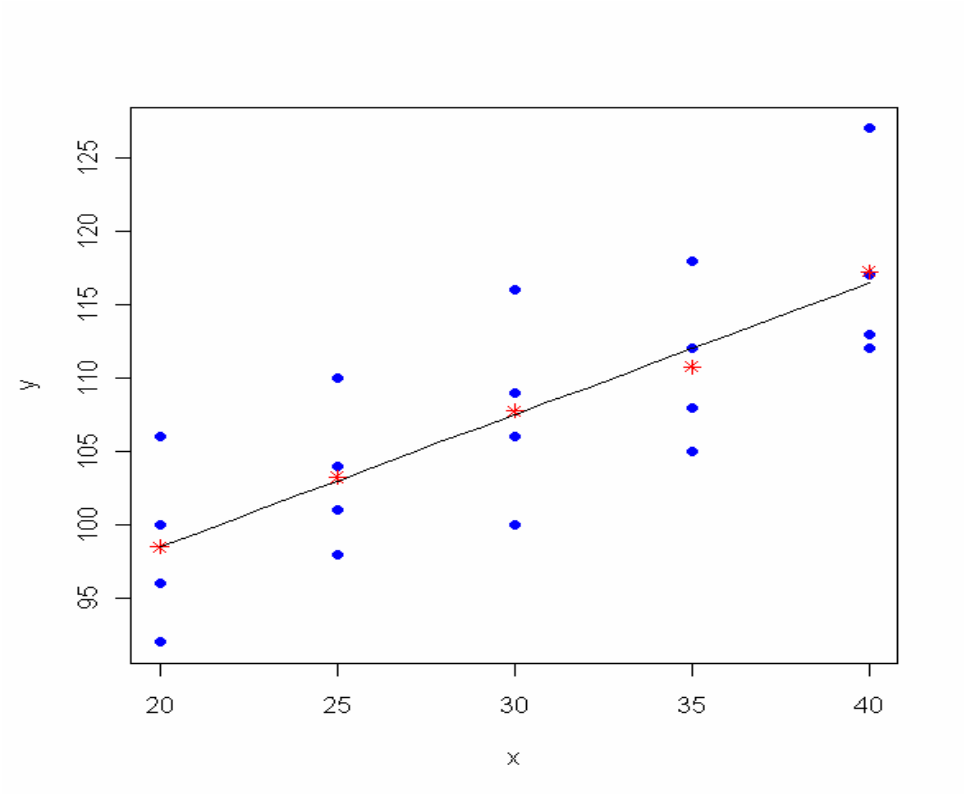
```
mod1<-lm(y~x)
anova(mod1)
```

```
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 810.00  810.00  25.897 7.662e-05 ***
Residuals 18 563.00   31.28
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

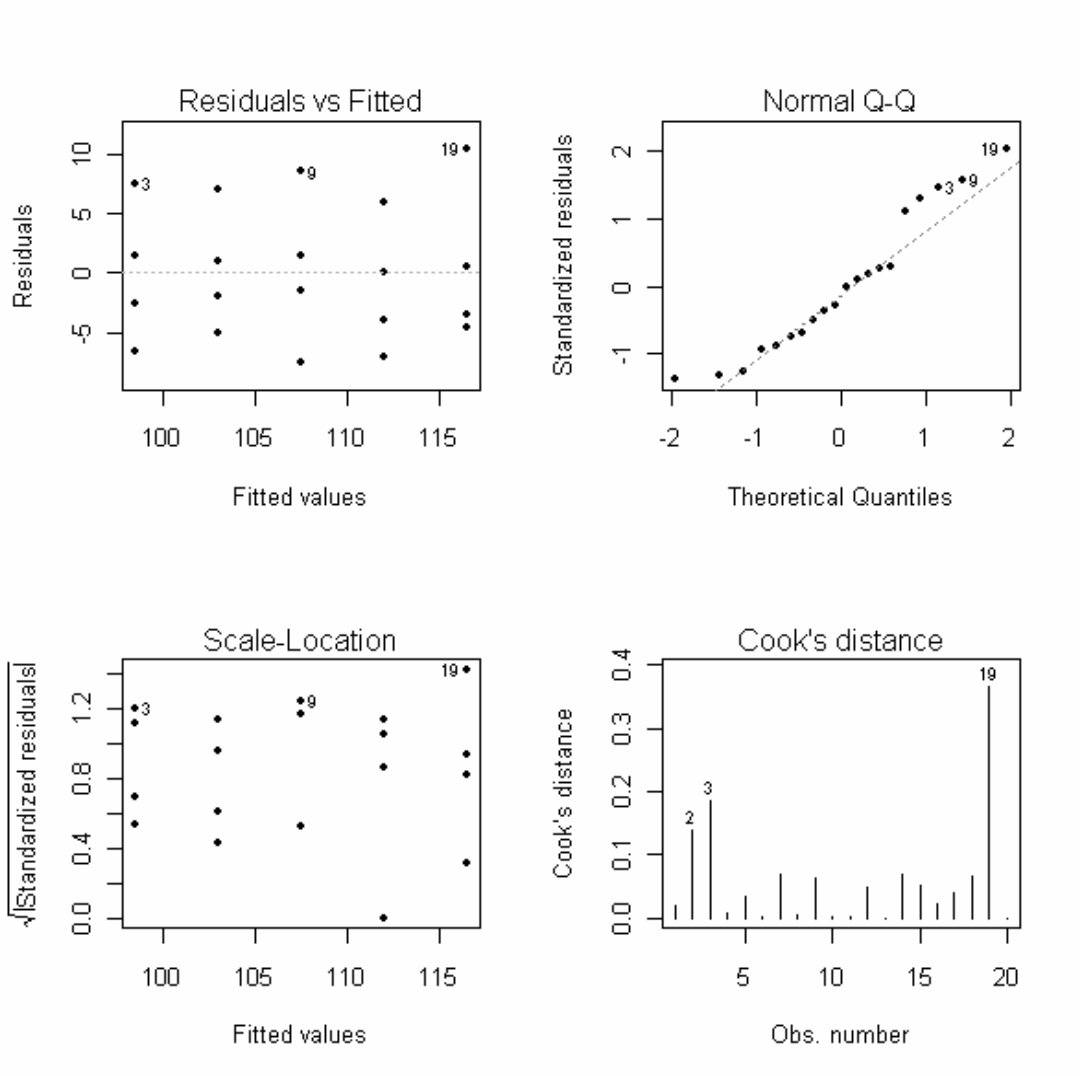
```
summary(mod1)
```

```
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-7.500 -4.125 -0.750  2.625 10.500

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80.5000     5.4510  14.768 1.67e-11 ***
x             0.9000     0.1769   5.089 7.66e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.593 on 18 degrees of freedom
Multiple R-squared: 0.5899,     Adjusted R-squared: 0.5672
F-statistic:  25.9 on 1 and 18 DF, p-value: 7.662e-05
```

```
x1<-20:40
ye<-80.5+0.9*x1
lines(x1,ye)
```




```
par(mfrow=c(2,2))
plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)
```




```
shapiro.test(mod1$resid)
```

```
Shapiro-Wilk normality test
W = 0.9376, p-value = 0.2161
```

Obs: Como para cada valor de x nesse exemplo há mais de um y observado (repetições), é possível utilizar o teste de Bartlett para testar a homogeneidade de variâncias.

bartlett.test(y~x)

```
        Bartlett test of homogeneity of variances

data:  y by x
Bartlett's K-squared = 0.2987, df = 4, p-value = 0.9899
```

predict(lm(y~x),interval="confidence", se.fit=T, level=0.95)

```
$fit
     fit       lwr      upr
1   98.5  93.94935 103.0507
2   98.5  93.94935 103.0507
3   98.5  93.94935 103.0507
4   98.5  93.94935 103.0507
5  103.0  99.78220 106.2178
6  103.0  99.78220 106.2178
7  103.0  99.78220 106.2178
8  103.0  99.78220 106.2178
9  107.5 104.87268 110.1273
10 107.5 104.87268 110.1273
11 107.5 104.87268 110.1273
12 107.5 104.87268 110.1273
13 112.0 108.78220 115.2178
14 112.0 108.78220 115.2178
15 112.0 108.78220 115.2178
16 112.0 108.78220 115.2178
17 116.5 111.94935 121.0507
18 116.5 111.94935 121.0507
19 116.5 111.94935 121.0507
20 116.5 111.94935 121.0507

$se.fit
```

```
 [1] 2.166026 2.166026 2.166026 2.166026 1.531611 1.531611 1.531611 1.531611
 [9] 1.250555 1.250555 1.250555 1.250555 1.531611 1.531611 1.531611 1.531611
[17] 2.166026 2.166026 2.166026 2.166026

$df
[1] 18

$residual.scale
[1] 5.592654
```

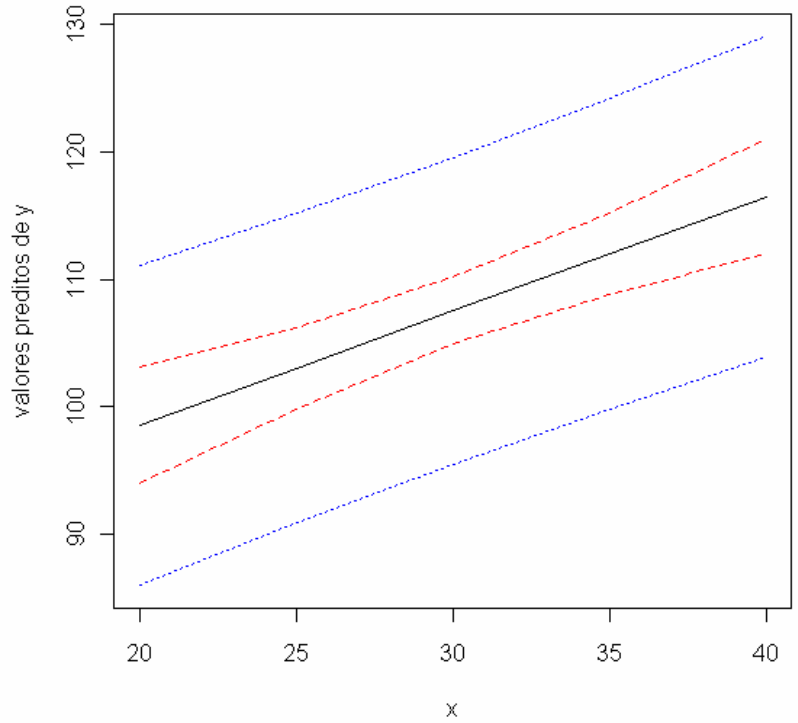
```
new <- data.frame(x = seq(24,28,1))
new
```

```
   x
1 24
2 25
3 26
4 27
5 28
```

```
predict(lm(y~x),new,interval="prediction", se.fit=T, level=0.95)
```

```
$fit
    fit      lwr      upr
1 102.1 89.85545 114.3445
2 103.0 90.81762 115.1824
3 103.9 91.76872 116.0313
4 104.8 92.70862 116.8914
5 105.7 93.63720 117.7628

$se.fit
       1        2        3        4        5
1.640088 1.531611 1.436779 1.358451 1.299615

$df
[1] 18

$residual.scale
[1] 5.592654
```

```
pred<- predict(lm(y~x),interval="confidence",  level=0.95)
pred1<- predict(lm(y~x),interval="prediction", level=0.95)
matplot(x,cbind(pred, pred1[,-1]), lty=c(1,2,2,3,3), col=c(1,2,2,4,4),
        type="l", ylab="valores preditos de y")
```



==========

EXEMPLO 2 Fonte: Montgomery e Peck (1992).

Y = tempo necessário para um entregador repor e executar pequenos serviços em máquinas automáticas de vendas de refrigerantes (em minutos)

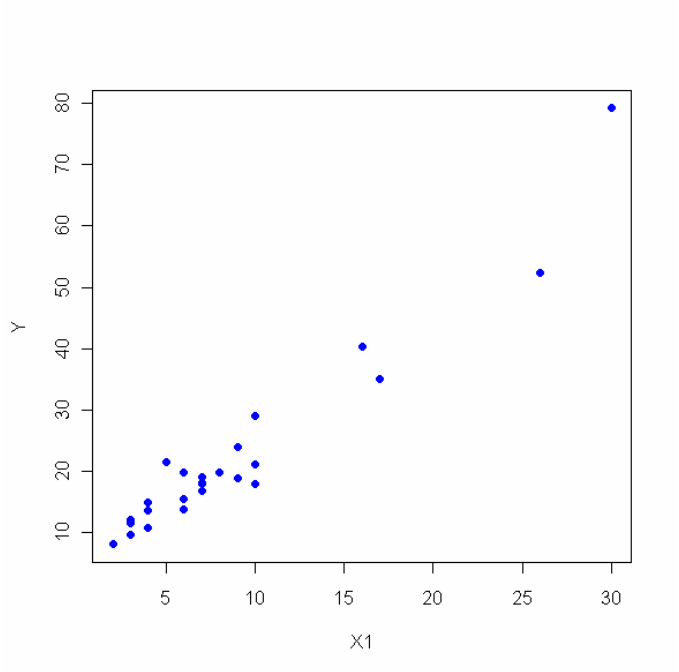
X = quantidade de volumes repostos nas máquinas (em unidades)

==========

```
ex2<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo2s.txt",h=T)
attach(ex2)
ex2
```

```
       Y X1
1  16.68  7
2  11.50  3
3  12.03  3
4  14.88  4
5  13.75  6
6  18.11  7
7   8.00  2
8  17.83  7
9  79.24 30
10 21.50  5
11 40.33 16
12 21.00 10
13 13.50  4
14 19.75  6
15 24.00  9
16 29.00 10
17 15.35  6
18 19.00  7
19  9.50  3
20 35.10 17
21 17.90 10
22 52.32 26
23 18.75  9
24 19.83  8
25 10.75  4
```

```
plot(X1, Y, pch=16, col=4)
```




```
cor(X1,Y)
[1] 0.9646146

cor.test(X1,Y)
        Pearson's product-moment correlation

data:  X1 and Y
t = 17.5455, df = 23, p-value = 8.216e-15
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9202275 0.9845031
sample estimates:
      cor
0.9646146

mod1<-lm(Y~X1)
```

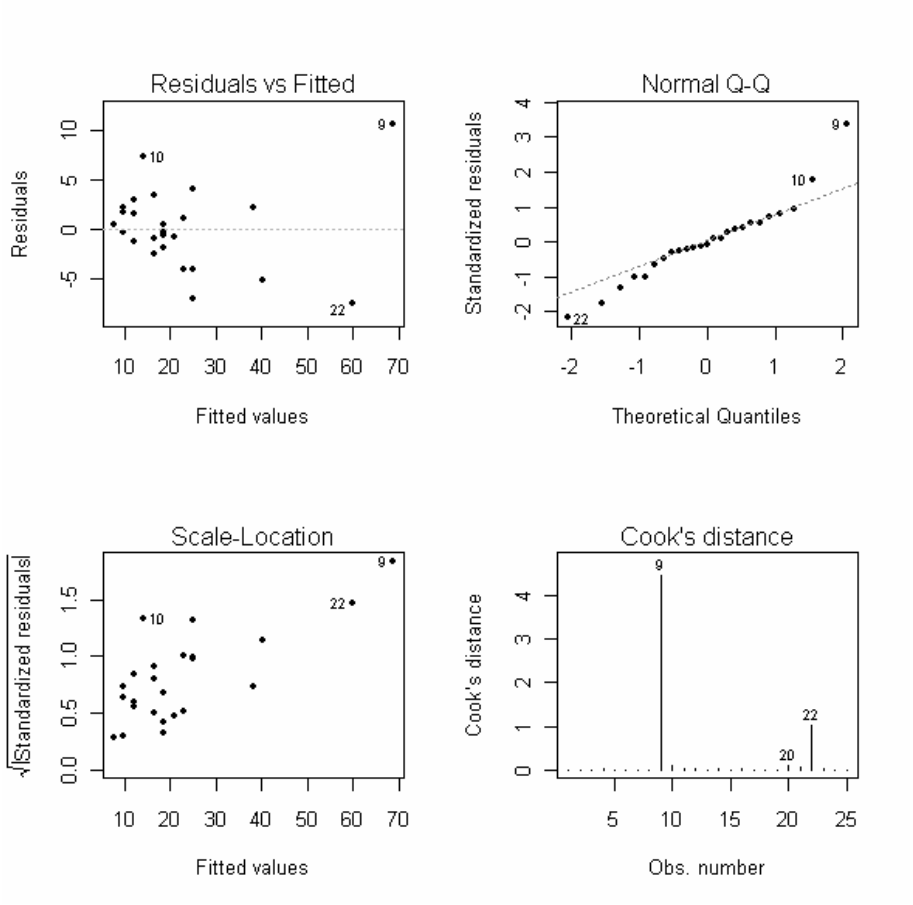
anova(mod1)

```
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
X1         1 5382.4  5382.4  307.85 8.22e-15 ***
Residuals 23  402.1    17.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

summary(mod1)

```
Residuals:
    Min      1Q  Median      3Q     Max
-7.5811 -1.8739 -0.3493  2.1807 10.6342

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.321      1.371   2.422   0.0237 *
X1             2.176      0.124  17.546 8.22e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 4.181 on 23 degrees of freedom
Multiple R-squared: 0.9305,     Adjusted R-squared: 0.9275
F-statistic: 307.8 on 1 and 23 DF,  p-value: 8.22e-15
```

par(mfrow=c(2,2))
plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)

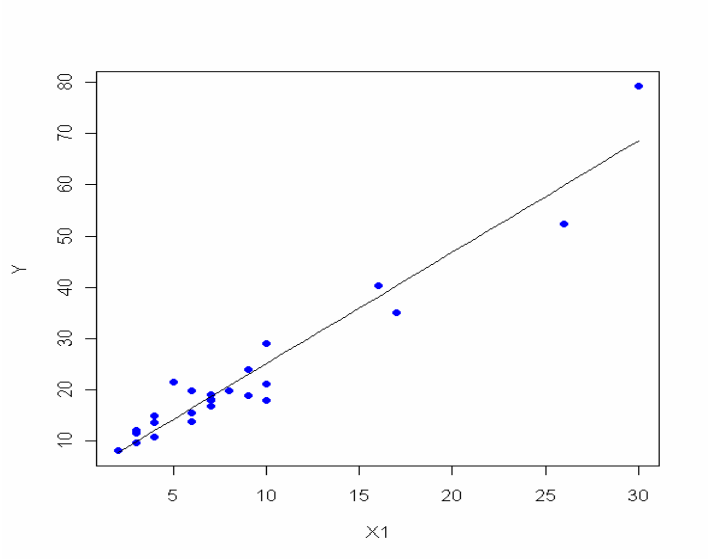


shapiro.test(mod1$resid)

```
        Shapiro-Wilk normality test

  data:  mod1$resid
  W = 0.9671, p-value = 0.5718
```

x1<-2:30

```
ye<-3.321+2.176*x1
lines(x1,ye)
```



==========

EXEMPLO 3

Y = tempo necessário para um comerciante estocar uma prateleira da mercearia com refrigerantes (em minutos)

X = quantidade da mercadoria (em unidades)

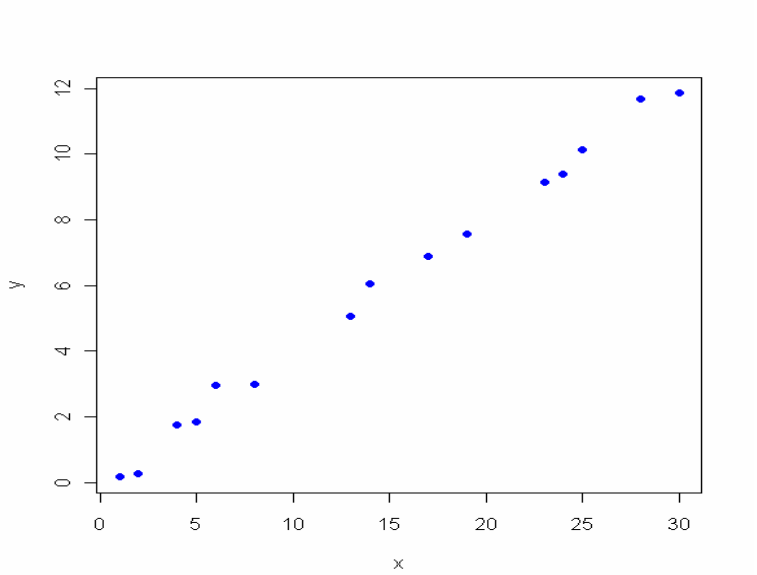
Fonte: Montgomery & Peck (1992).

==========

```
ex3<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo3s.txt",h=T)
attach(ex3)
       y  x
1  10.15 25
2   2.96  6
3   3.00  8
4   6.88 17
5   0.28  2
```

```
6   5.06 13
7   9.14 23
8  11.86 30
9  11.69 28
10  6.04 14
11  7.57 19
12  1.74  4
13  9.38 24
14  0.16  1
15  1.84  5
```

```
plot(x,y, pch=16, col=4)
```




```
cor(x,y)
[1] 0.9973597
```

```
cor.test(x,y)

        Pearson's product-moment correlation
```

```
data:  x and y
t = 49.519, df = 13, p-value = 4.441e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9918363 0.9991477
sample estimates:
      cor
0.9973597
```

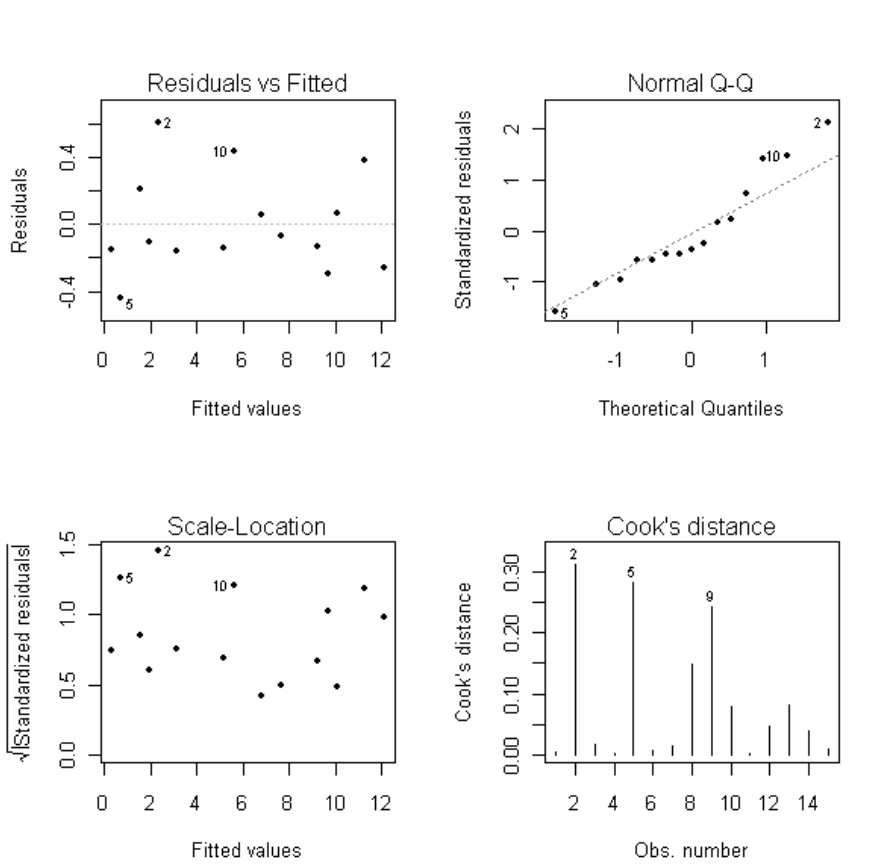
```
mod1<-lm(y~x)
anova(mod1)
```

```
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 228.318 228.318  2452.1 3.399e-16 ***
Residuals 13   1.210   0.093
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
summary(mod1)
```

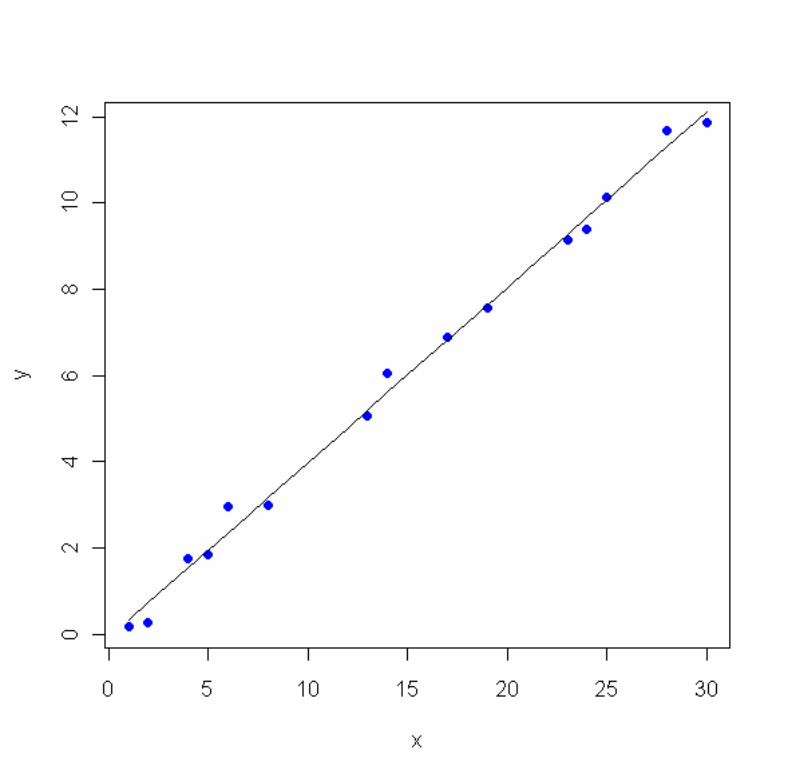
```
Residuals:
    Min      1Q  Median      3Q     Max
-0.4405 -0.1582 -0.1018  0.1357  0.6111

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.093756   0.143577  -0.653    0.525
x            0.407107   0.008221  49.519  3.4e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3051 on 13 degrees of freedom
Multiple R-squared: 0.9947,     Adjusted R-squared: 0.9943
F-statistic:  2452 on 1 and 13 DF,  p-value: 3.399e-16
```

```
par(mfrow=c(2,2))
plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)
```




```
shapiro.test(mod1$resid)

  Shapiro-Wilk normality test
  W = 0.9321, p-value = 0.2937

plot(x,y, pch=16, col=4)
x1<-1:30
ye<- -0.093756+0.407107*x1
```

```
lines(x1,ye)
```



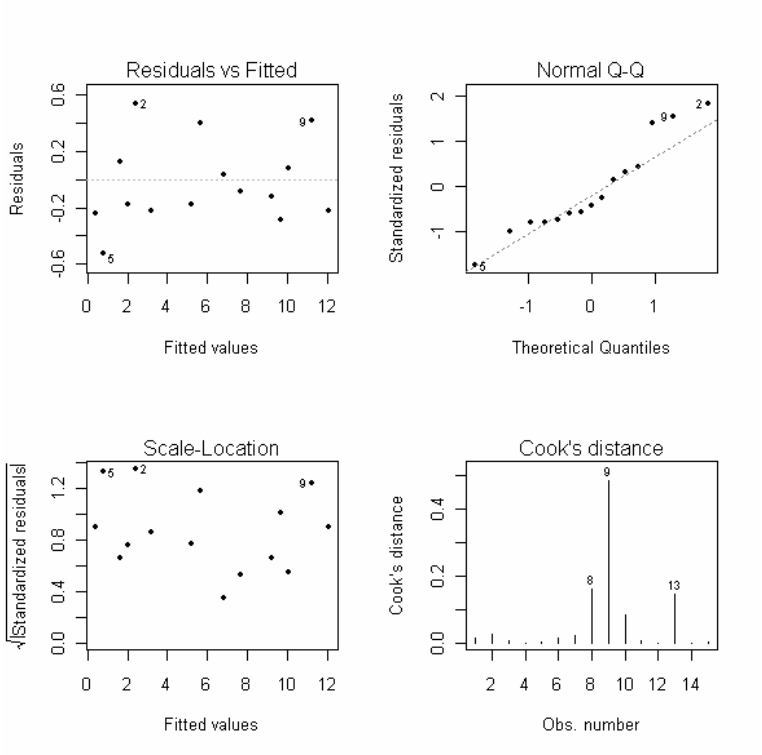

```
============================================================
# limite inferior da variação de x muito próxima de zero
# e faz sentido que quando x = 0 se tenha y = 0
============================================================

mod2<-lm(y~-1 + x)    # modelo sem intercepto
summary(mod2)

Call:
lm(formula = y ~ -1 + x)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-0.5252 -0.2198 -0.1202  0.1070  0.5443

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x 0.402619   0.004418   91.13   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2988 on 14 degrees of freedom
Multiple R-squared: 0.9983,     Adjusted R-squared: 0.9982
F-statistic:  8305 on 1 and 14 DF,  p-value: < 2.2e-16
```

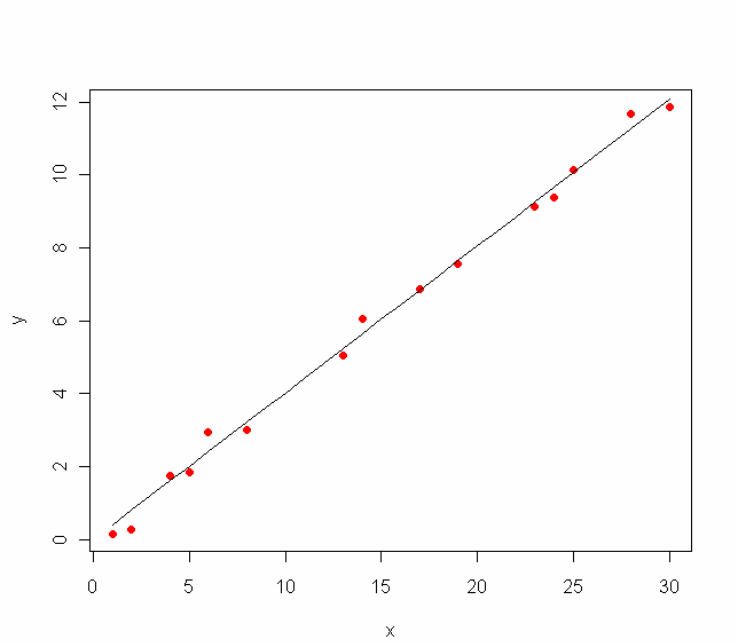
anova(mod2)

```
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 741.62  741.62  8305.2 < 2.2e-16 ***
Residuals 14   1.25    0.09
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

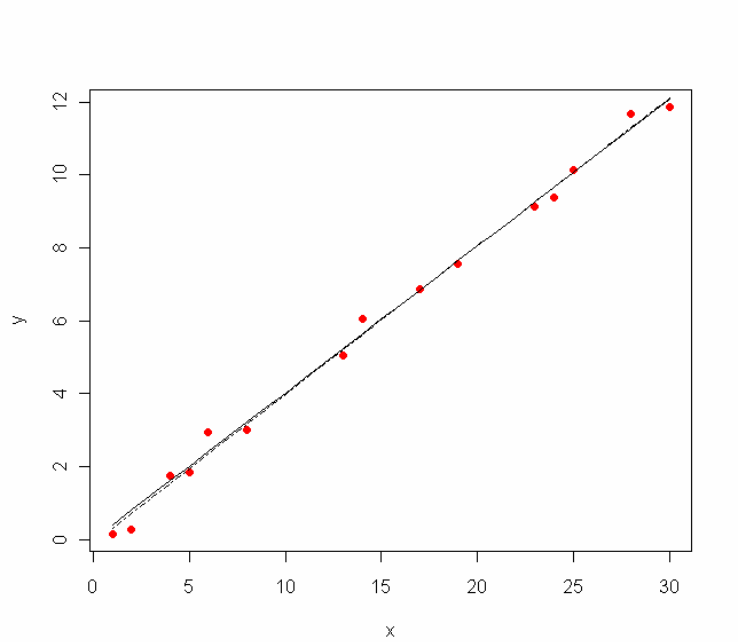
par(mfrow=c(2,2))
plot(mod2, which=c(1:4), add.smooth=FALSE, pch=20)




```
shapiro.test(mod2$resid)

        Shapiro-Wilk normality test

data:  mod2$resid
W = 0.929, p-value = 0.2639

par(mfrow=c(1,1))
plot(x,y, pch=16, col=2)
x1<-min(x):max(x)
ye<-mod2$coef[1]*x1
lines(x1,ye)
```




```
ye1<-mod1$coef[1]+mod1$coef[2]*x1
lines(x1,ye1, lty=4)
```



==========

EXEMPLO 4

Y = Renda média mensal sobre vendas de comida em restaurantes (em mil dólares)

X = Despesas anual com propagandas (em mil dólares)

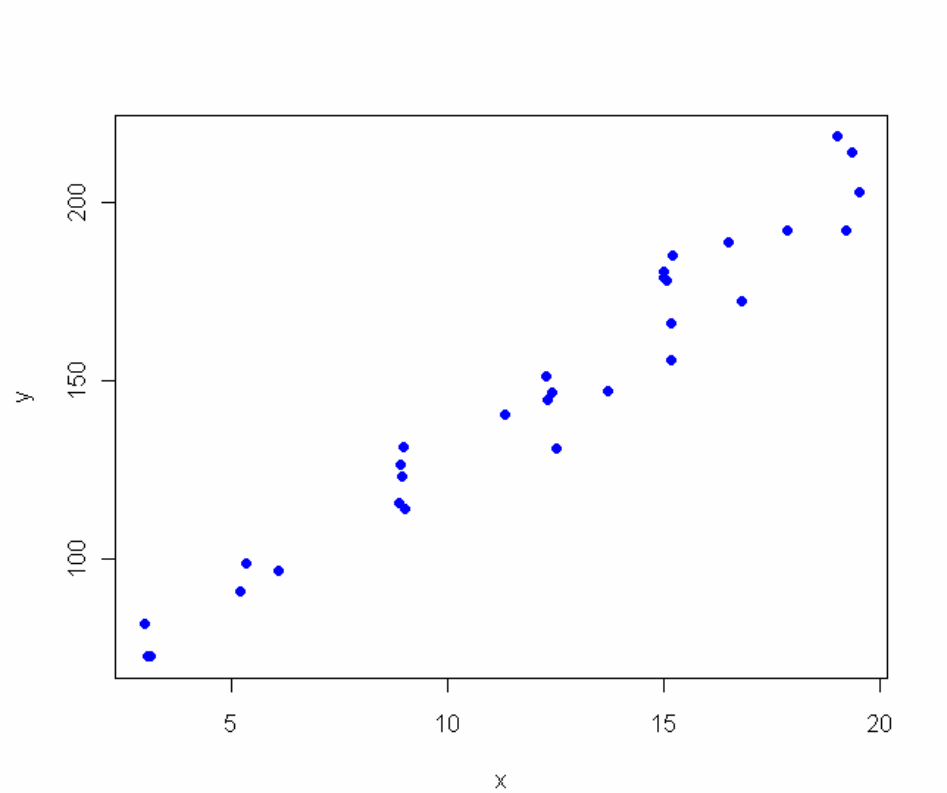
Fonte: Montgomery e Peck (1992)

==========

```
 rm(list = ls())
 ex4<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo4s.txt",h=T)
 attach(ex4)
 ex4
         y      x
1   81.464  3.000
2   72.661  3.150
3   72.344  3.085
4   90.743  5.225
```

```
5   98.588  5.350
6   96.507  6.090
7  126.574  8.925
8  114.133  9.015
9  115.814  8.885
10 123.181  8.950
11 131.434  9.000
12 140.564 11.345
13 151.352 12.275
14 146.926 12.400
15 130.963 12.525
16 144.630 12.310
17 147.041 13.700
18 179.021 15.000
19 166.200 15.175
20 180.732 14.995
21 178.187 15.050
22 185.304 15.200
23 155.931 15.150
24 172.579 16.800
25 188.851 16.500
26 192.424 17.830
27 203.112 19.500
28 192.482 19.200
29 218.715 19.000
30 214.317 19.350
```

```
plot(x,y, pch=16, col=4)
```



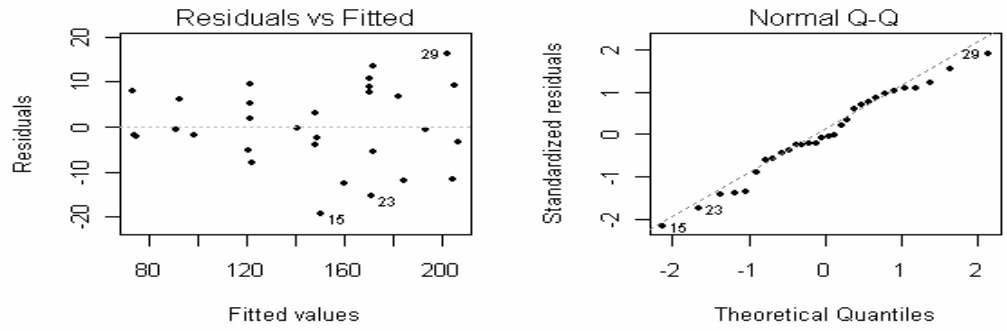
cor(x,y)
[1] 0.9777232

cor.test(x,y)

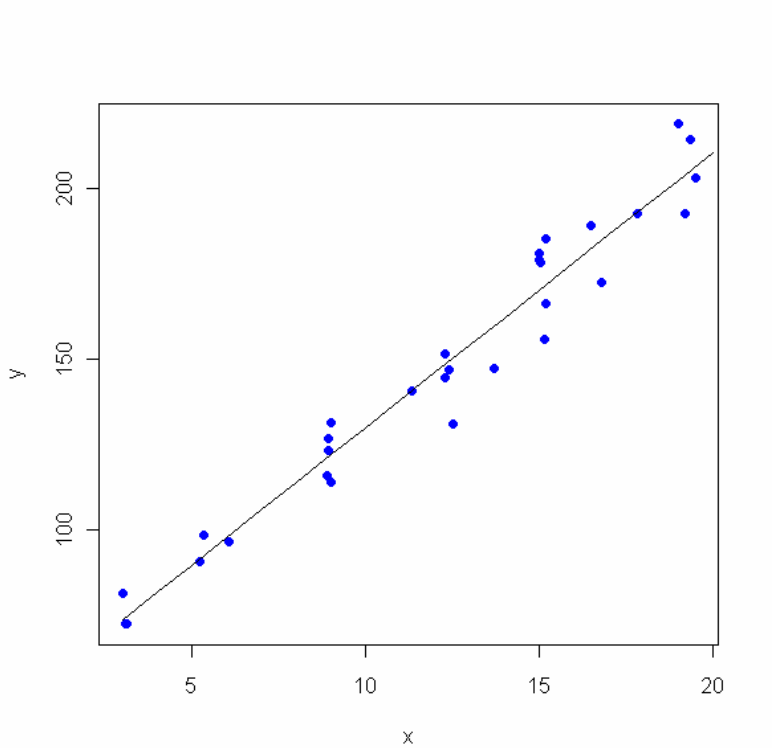
```
        Pearson's product-moment correlation

data:  x and y
t = 24.6482, df = 28, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
```

```
 0.9532193 0.9894611
sample estimates:
      cor
0.9777232

mod1<-lm(y~x)
anova(mod1)

Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1  49200   49200  607.53 < 2.2e-16 ***
Residuals 28   2268      81
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 summary(mod1)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-19.2871  -4.8273  -0.6383   7.3630  16.3512

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.4434     4.2889   11.53 3.81e-12 ***
x             8.0484     0.3265   24.65  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.999 on 28 degrees of freedom
Multiple R-squared: 0.9559,     Adjusted R-squared: 0.9544
F-statistic: 607.5 on 1 and 28 DF,  p-value: < 2.2e-16

par(mfrow=c(2,2))
plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)
shapiro.test(mod1$resid)
```

```
        Shapiro-Wilk normality test

data:  mod1$resid
W = 0.9774, p-value = 0.7523
```



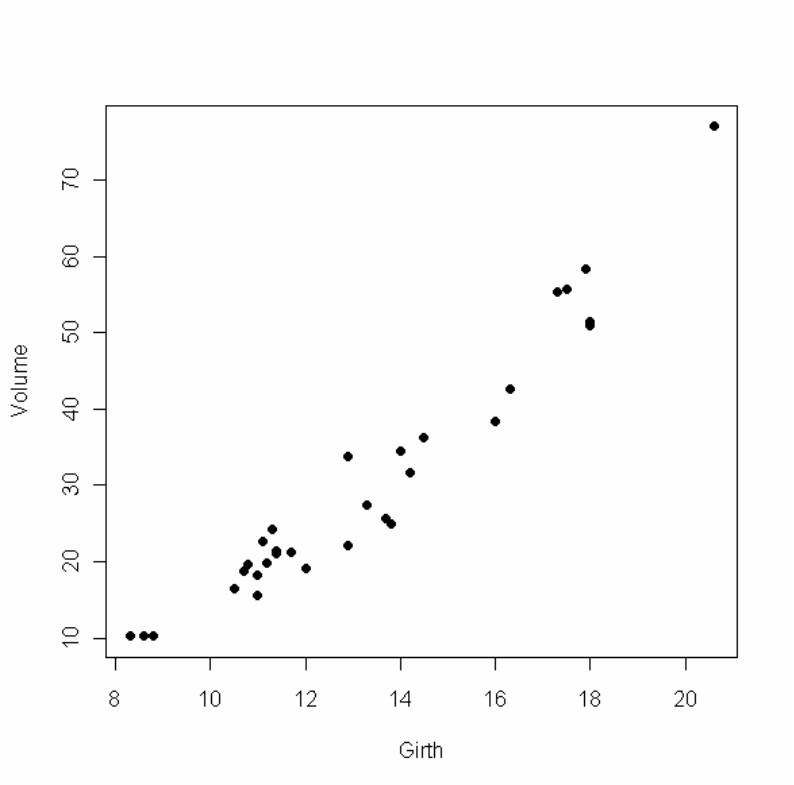

```
x1<-3:20
ye<-49.4434+8.0484*x1
plot(x,y, pch=16, col=4)
lines(x1,ye,type="l")
```




```
================================================================================
EXEMPLO 5: Dados: trees (Fonte: Disponível no R)
Y = diâmetro de árvores (em polegadas)
X = volume de madeira (em pés cúbicos)
================================================================================

data(trees)
help(trees)
attach(trees)
trees

  Girth Height Volume
```

```
1    8.3     70   10.3
2    8.6     65   10.3
3    8.8     63   10.2
4   10.5     72   16.4
5   10.7     81   18.8
6   10.8     83   19.7
7   11.0     66   15.6
8   11.0     75   18.2
9   11.1     80   22.6
10  11.2     75   19.9
11  11.3     79   24.2
12  11.4     76   21.0
13  11.4     76   21.4
14  11.7     69   21.3
15  12.0     75   19.1
16  12.9     74   22.2
17  12.9     85   33.8
18  13.3     86   27.4
19  13.7     71   25.7
20  13.8     64   24.9
21  14.0     78   34.5
22  14.2     80   31.7
23  14.5     74   36.3
24  16.0     72   38.3
25  16.3     77   42.6
26  17.3     81   55.4
27  17.5     82   55.7
28  17.9     80   58.3
29  18.0     80   51.5
30  18.0     80   51.0
31  20.6     87   77.0
```

```
plot(Girth,Volume)
```




```
cor(Girth,Volume)
[1] 0.9671194

cor.test(Girth,Volume)

        Pearson's product-moment correlation

data:  Girth and Volume
t = 20.4783, df = 29, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
```

```
 0.9322519 0.9841887
sample estimates:
      cor
0.9671194
```

```
fit1<-lm(Volume~Girth)
summary(fit1)
```

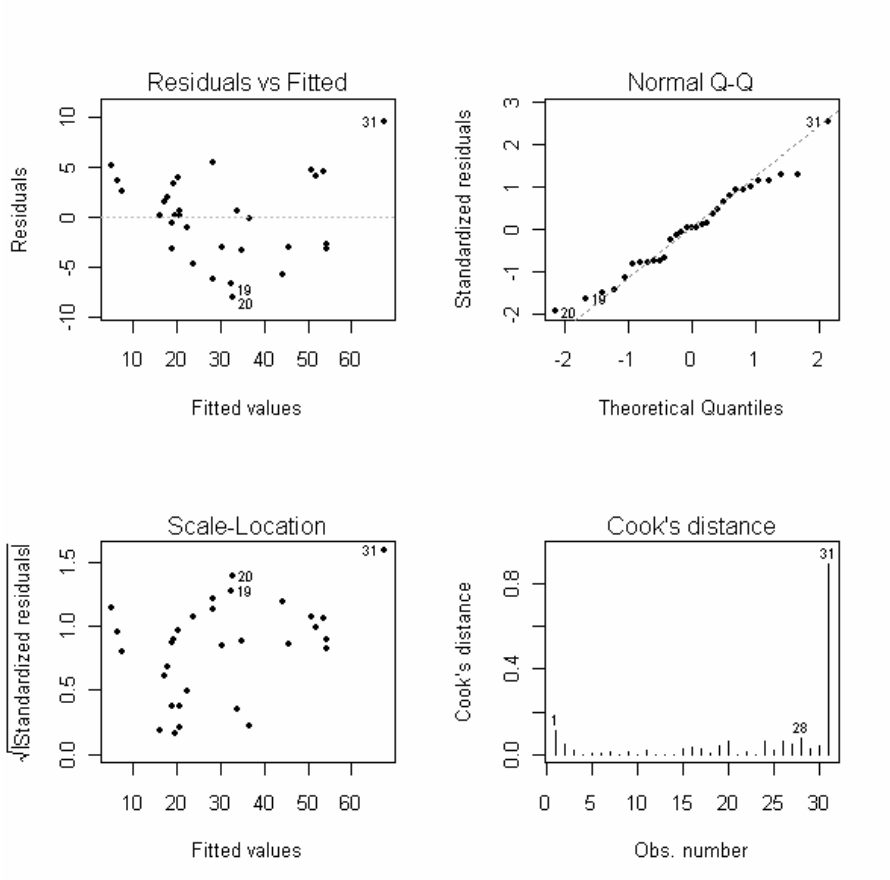
```
Call:
lm(formula = Volume ~ Girth)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0654 -3.1067  0.1520  3.4948  9.5868

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -36.9435     3.3651  -10.98 7.62e-12 ***
Girth         5.0659     0.2474   20.48  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.252 on 29 degrees of freedom
Multiple R-squared: 0.9353,     Adjusted R-squared: 0.9331
F-statistic: 419.4 on 1 and 29 DF,  p-value: < 2.2e-16
```

```
anova(fit1)
```

```
Analysis of Variance Table
Response: Volume
          Df Sum Sq Mean Sq F value    Pr(>F)
Girth      1 7581.8  7581.8  419.36 < 2.2e-16 ***
Residuals 29  524.3    18.1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>  shapiro.test(fit1$res)

        Shapiro-Wilk normality test
```

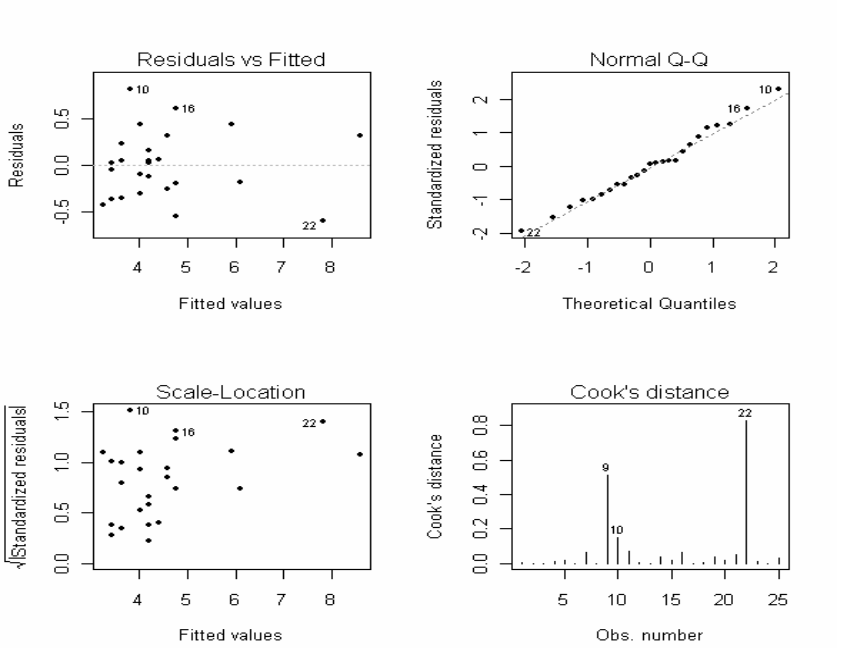
```
data:  fit1$res
W = 0.9789, p-value = 0.7811
```

```
par(mfrow=c(2,2))
plot(fit1, which=c(1:4), add.smooth=FALSE, pch=20)
```

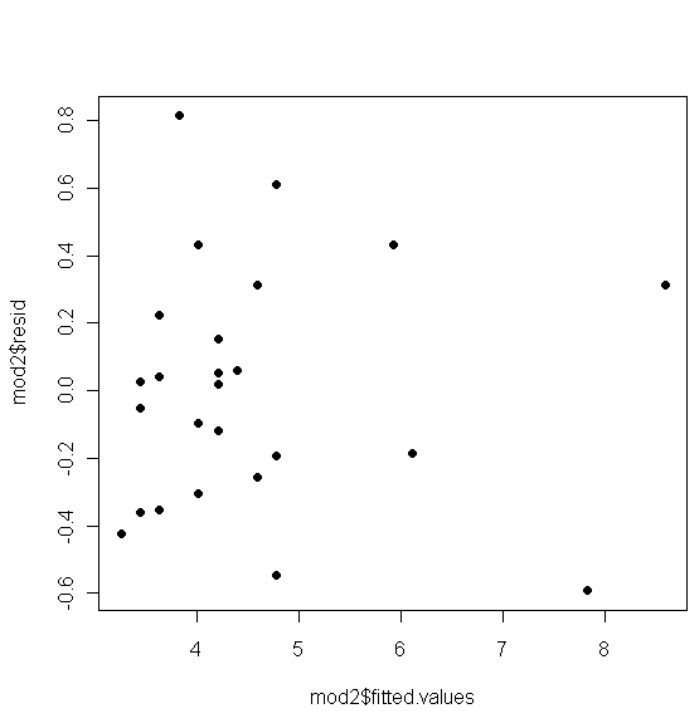


Como podemos tentar melhorar esse modelo?

```
################################################
 TRANSFORMAÇÕES EM Y
################################################

 Exemplo2 – Transformação Raiz Quadrado em Y

 ex2<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo2s.txt",h=T)
 attach(ex2)
 mod2<-lm(sqrt(Y)~X1)
 par(mfrow=c(2,2))
 plot(mod2, which=c(1:4), add.smooth=FALSE, pch=20)
```




```
 plot(mod2$fitted.values, mod2$resid, pch=16)
```



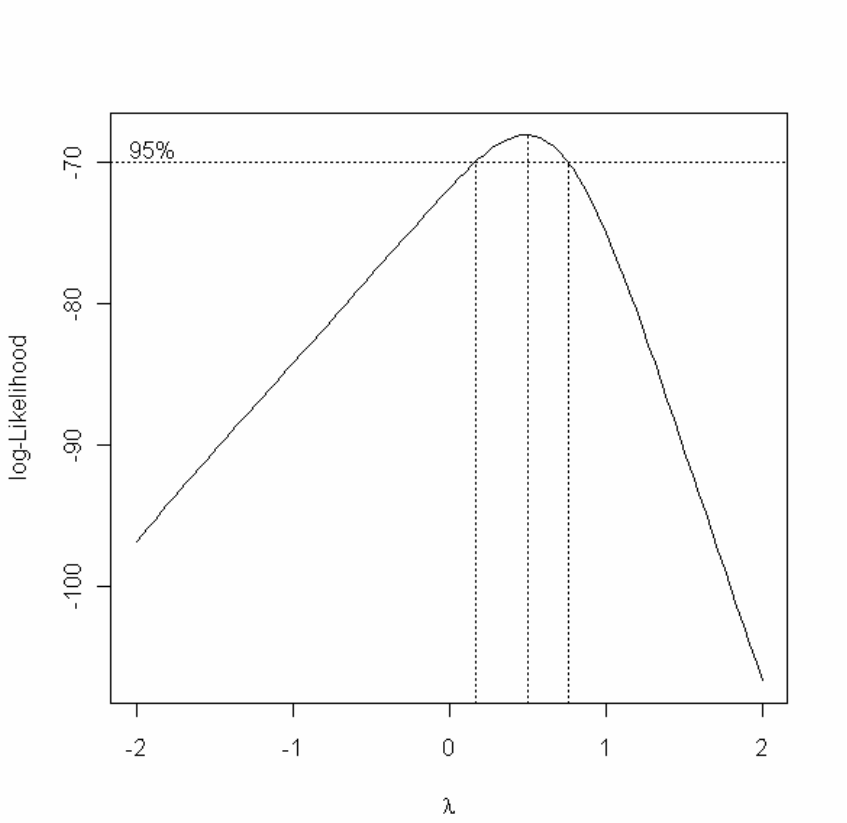
summary(mod2)

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.59277 -0.25560  0.01804  0.22476  0.81350

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.87028    0.11886   24.15  < 2e-16 ***
X1           0.19061    0.01075   17.73  6.6e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

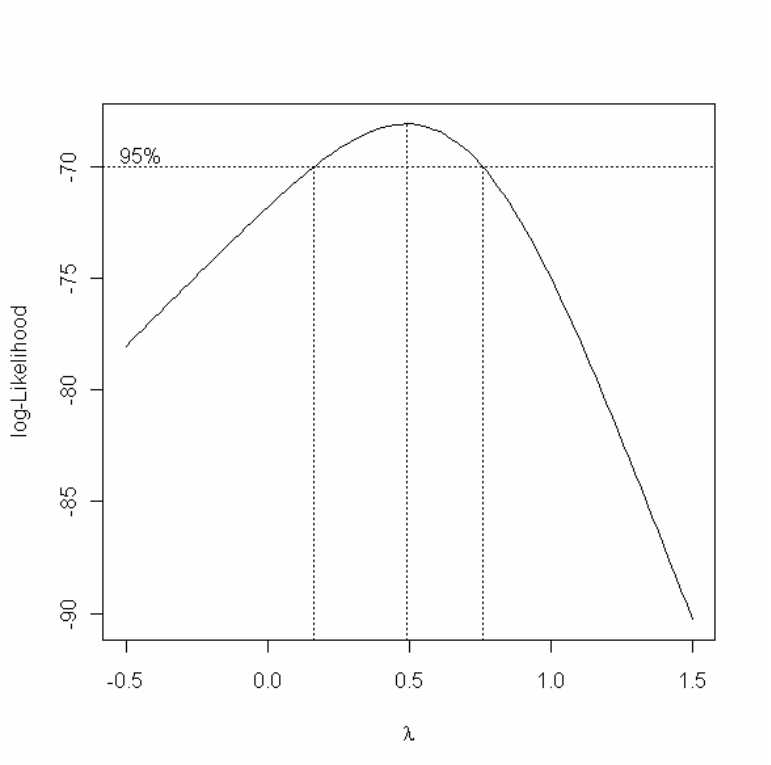
```
Residual standard error: 0.3625 on 23 degrees of freedom
Multiple R-squared: 0.9318,     Adjusted R-squared: 0.9288
F-statistic: 314.2 on 1 and 23 DF,  p-value: 6.594e-15
```

```
 shapiro.test(mod2$resid)

        Shapiro-Wilk normality test

data:  mod2$resid
W = 0.9794, p-value = 0.8725
```

```
 require(MASS)
 boxcox(Y~X1,data=ex2,plotit=T)
```




```
boxcox(Y~X1,data=ex2,lam=seq(-0.5,1.5,1/10))
```

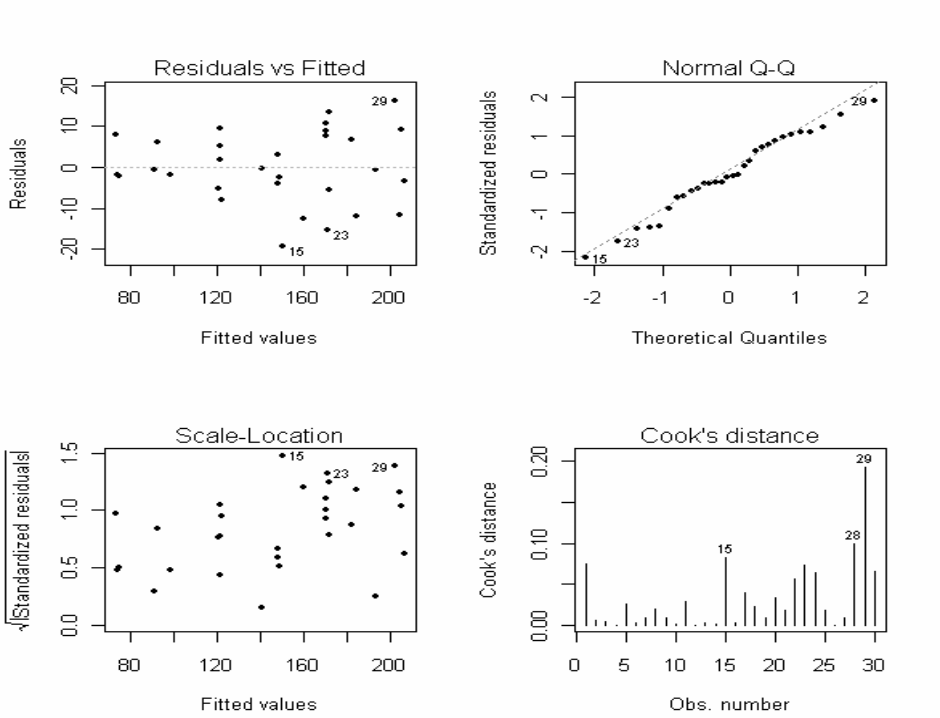


boxcox(Y~X1,data=ex2,plotit=F)

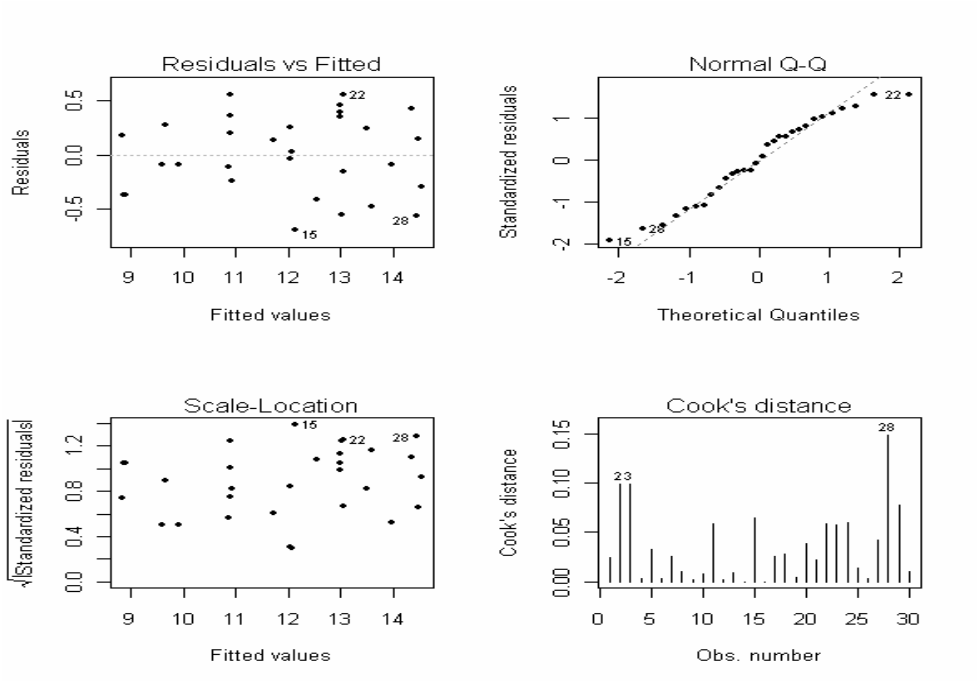
```
$x
 [1] -2.0 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6
[16] -0.5 -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
[31]  1.0  1.1  1.2  1.3  1.4  1.5  1.6  1.7  1.8  1.9  2.0

$y
 [1]  -96.83566  -95.53286  -94.24200  -92.96237  -91.69314  -90.43331
 [7]  -89.18172  -87.93706  -86.69782  -85.46232  -84.22875  -82.99519
[13]  -81.75969  -80.52043  -79.27594  -78.02549  -76.76964  -75.51119
[19]  -74.25647  -73.01738  -71.81411  -70.67886  -69.66004  -68.82601
```

```
[25]  -68.26562  -68.08184  -68.37491  -69.21697  -70.62856  -72.57098
[31]  -74.95981  -77.69050  -80.66194  -83.79002  -87.01158  -90.28253
[37]  -93.57378  -96.86716 -100.15195 -103.42240 -106.67591
```

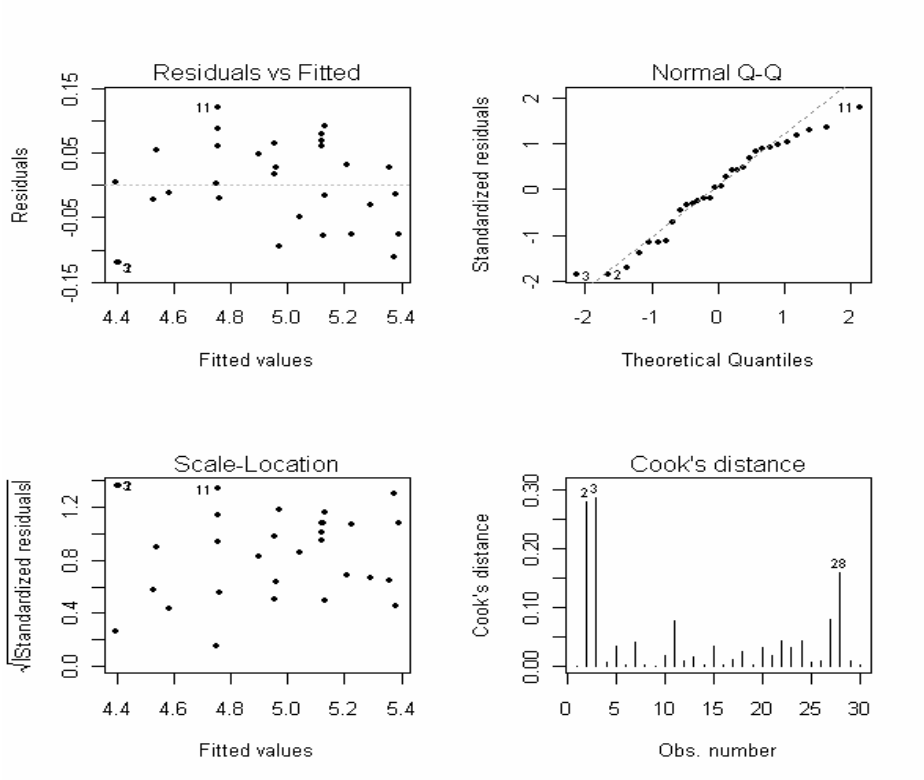
```
#################################################################
Exemplo 4 – Transformações em Y
#################################################################

rm(list = ls())
ex4<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo4s.txt",h=T)
attach(ex4)

mod1<-lm(y~x)
par(mfrow=c(2,2))
plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)
```



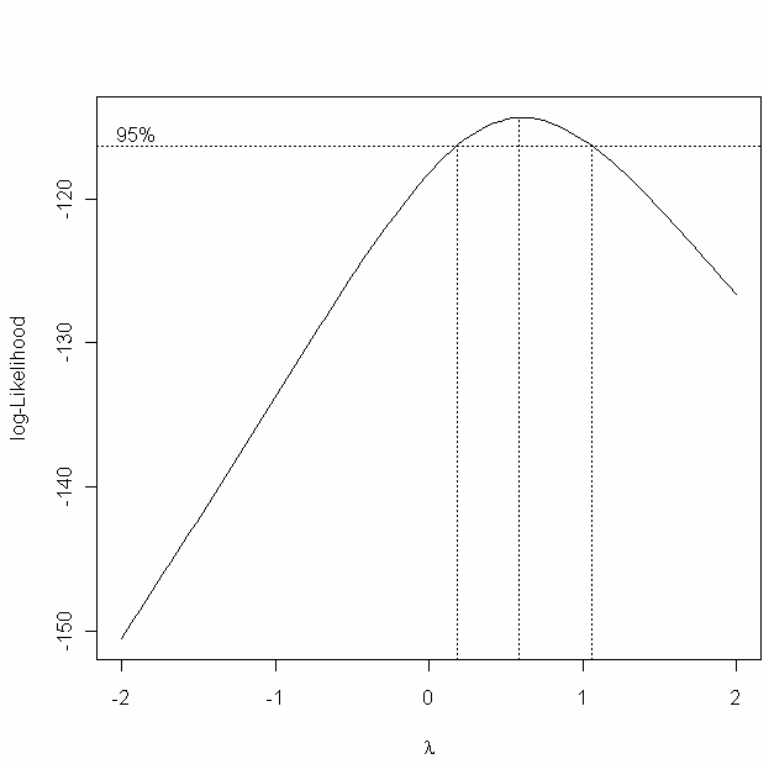

```
mod2<-lm(sqrt(y)~x)
par(mfrow=c(2,2))
plot(mod2, which=c(1:4), add.smooth=FALSE, pch=20)
```



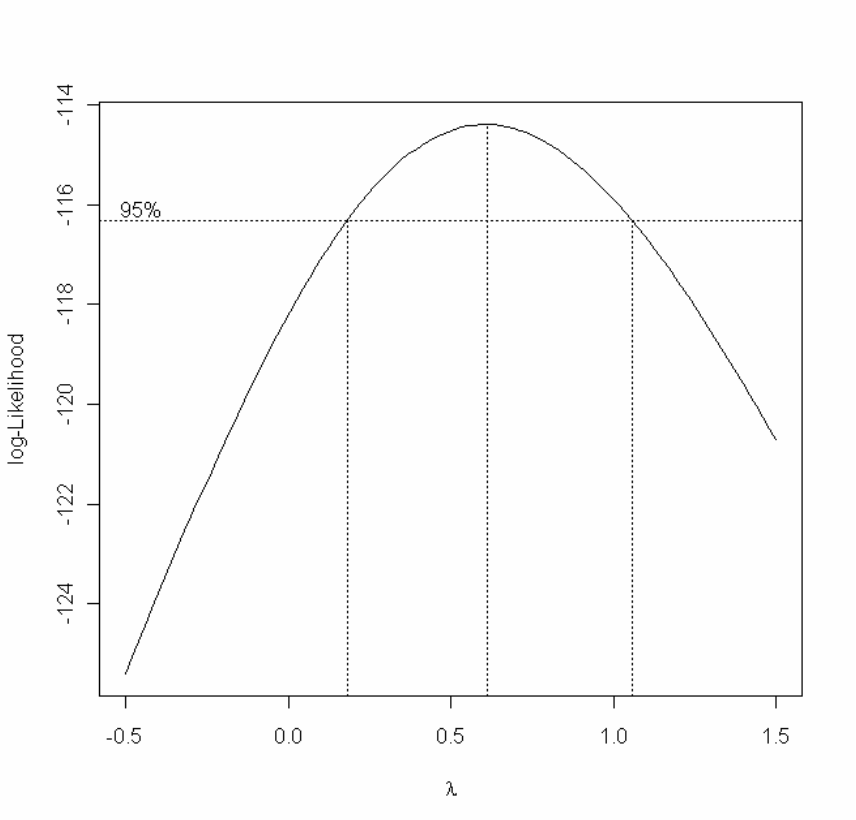

```
mod3<-lm(log(y)~x)
par(mfrow=c(2,2))
plot(mod3, which=c(1:4), add.smooth=FALSE, pch=20)
```



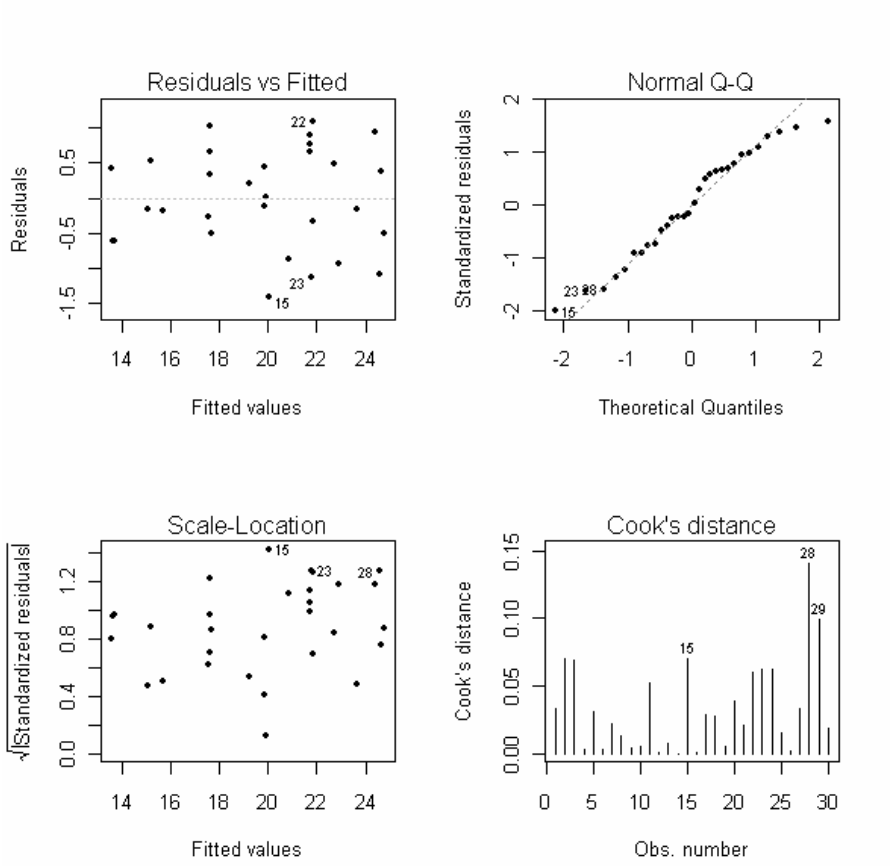

```
par(mfrow=c(1,1))
require(MASS)
boxcox(y~x,data=ex4,plotit=T)
```




```
boxcox(y~x,data=ex4,lam=seq(-0.5,1.5,1/10))
```



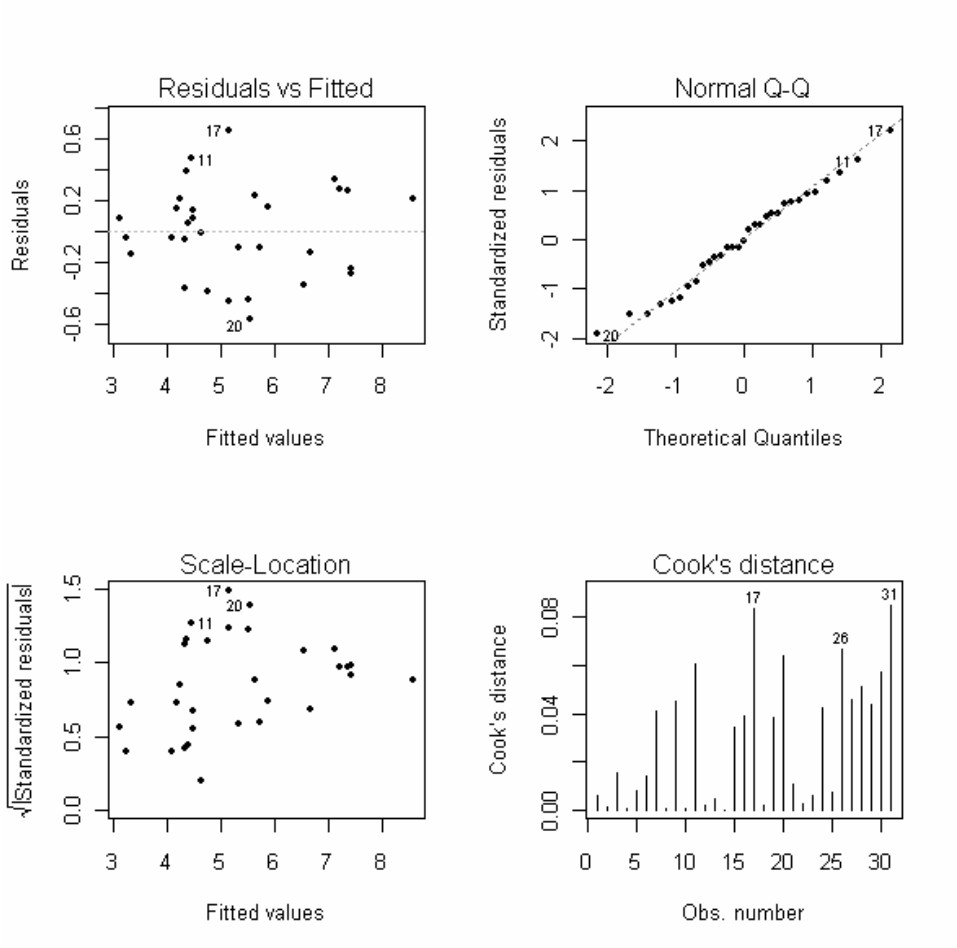

```
par(mfrow=c(2,2))
mod4<-lm((y^0.6)~x)
plot(mod4, which=c(1:4), add.smooth=FALSE, pch=20)
```



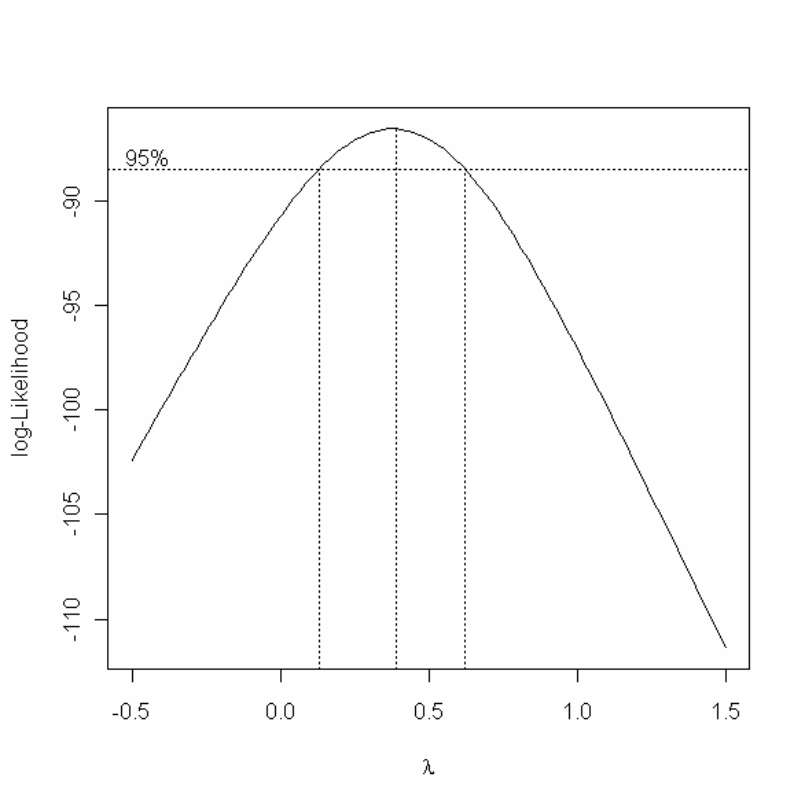

```
shapiro.test(mod1$resid)
 W = 0.9774, p-value = 0.7523

shapiro.test(mod2$resid)
 W = 0.9612, p-value = 0.3327

shapiro.test(mod3$resid)
 W = 0.9593, p-value = 0.2969
```

```
 shapiro.test(mod4$resid)
  W = 0.9643, p-value = 0.3971


#######################################################
Exemplo 5 – Transformações em Y
#######################################################

 data(trees)
 attach(trees)

 fit2<-lm(sqrt(Volume)~Girth)
 summary(fit2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.56640 -0.19429 -0.01169  0.20934  0.65575

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.55183    0.23719  -2.327   0.0272 *
Girth        0.44262    0.01744  25.385   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2997 on 29 degrees of freedom
Multiple R-squared: 0.9569,     Adjusted R-squared: 0.9555
F-statistic: 644.4 on 1 and 29 DF,  p-value: < 2.2e-16

 anova(fit2)

Analysis of Variance Table
Response: sqrt(Volume)
          Df Sum Sq Mean Sq F value    Pr(>F)
Girth      1 57.881  57.881  644.42 < 2.2e-16 ***
Residuals 29  2.605   0.090
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
 shapiro.test(fit2$res)

        Shapiro-Wilk normality test
data:  fit2$res
W = 0.9858, p-value = 0.9457

 par(mfrow=c(2,2))
 plot(fit2, which=c(1:4), add.smooth=FALSE, pch=20)
```




```
require(MASS)
boxcox(Volume~Girth,data=trees,plotit=T)
boxcox(Volume~Girth,data=trees,lam=seq(-0.5,1.5,1/10))
```

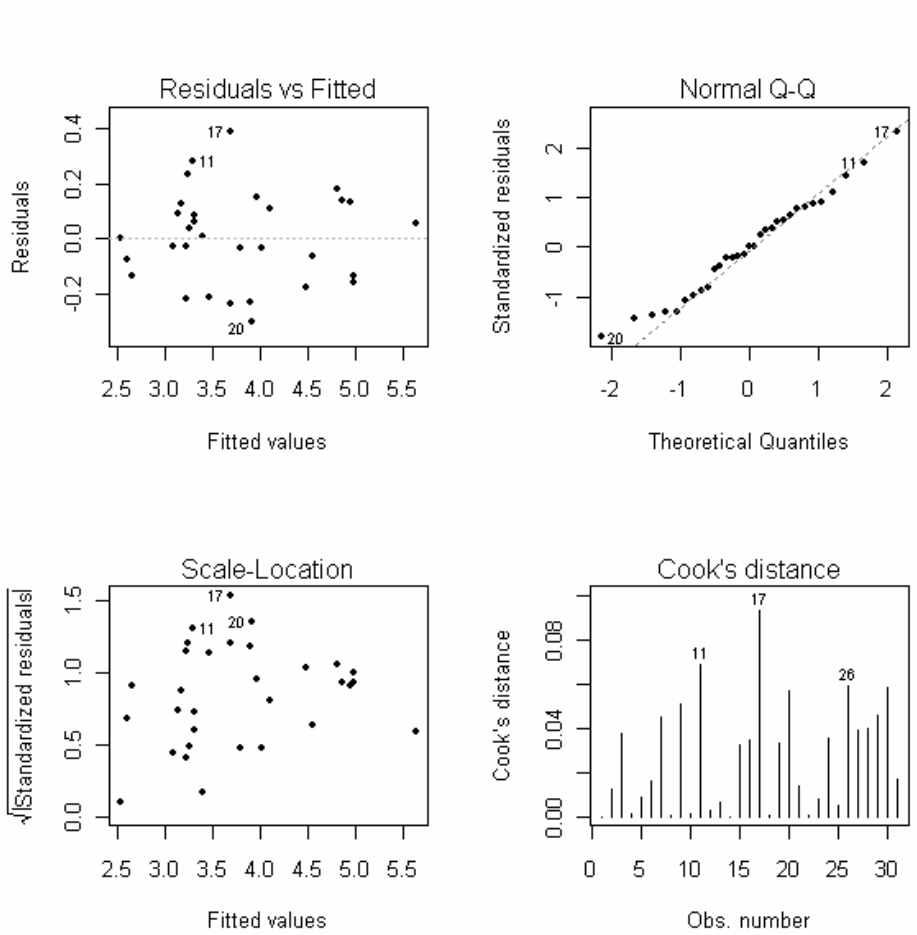



```
bc<-boxcox(Volume~Girth,data=trees,plotit=F)
cbind(bc$x,bc$y)
```

```
     [,1]       [,2]
[1,] -2.0 -137.69639
[2,] -1.9 -135.43059
[3,] -1.8 -133.16051
[4,] -1.7 -130.88450
[5,] -1.6 -128.60076
[6,] -1.5 -126.30727
[7,] -1.4 -124.00186
[8,] -1.3 -121.68223
```

```
 [9,] -1.2 -119.34600
[10,] -1.1 -116.99079
[11,] -1.0 -114.61444
[12,] -0.9 -112.21525
[13,] -0.8 -109.79244
[14,] -0.7 -107.34678
[15,] -0.6 -104.88164
[16,] -0.5 -102.40450
[17,] -0.4  -99.92915
[18,] -0.3  -97.47883
[19,] -0.2  -95.09024
[20,] -0.1  -92.81840
[21,]  0.0  -90.74102
[22,]  0.1  -88.95991
[23,]  0.2  -87.59521
[24,]  0.3  -86.76836
[25,]  0.4  -86.57447
[26,]  0.5  -87.05365
[27,]  0.6  -88.17730
[28,]  0.7  -89.85812
[29,]  0.8  -91.97739
[30,]  0.9  -94.41372
[31,]  1.0  -97.06207
[32,]  1.1  -99.84112
[33,]  1.2 -102.69275
[34,]  1.3 -105.57791
[35,]  1.4 -108.47175
[36,]  1.5 -111.35941
[37,]  1.6 -114.23277
[38,]  1.7 -117.08805
[39,]  1.8 -119.92427
[40,]  1.9 -122.74207
[41,]  2.0 -125.54309
```

```
fit3<-lm((Volume^0.4)~Girth)
par(mfrow=c(2,2))
plot(fit3, which=c(1:4), add.smooth=FALSE, pch=20)
```



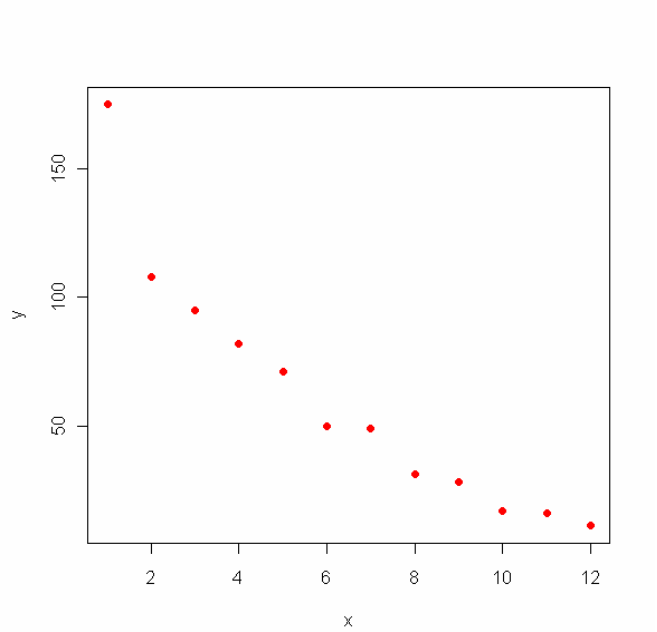

```
#######################################################
Exemplo 6 – Linearização
Y = número médio de bactérias sobreviventes
    em um produto alimentício enlatado
X = minutos de exposição a 300º F
Fonte: Montgomery & Peck (1992).
```

```
#######################################################
```

```
y<-c(175,108,95,82,71,50,49,31,28,17,16,11)
x<-1:12
cbind(y,x)
```

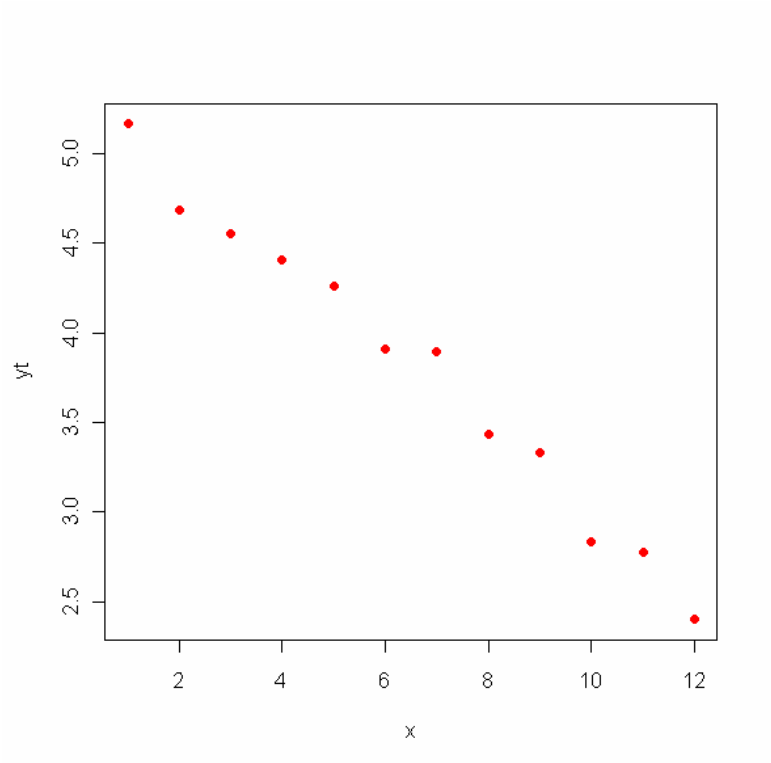
```
       y  x
 [1,] 175  1
 [2,] 108  2
 [3,]  95  3
 [4,]  82  4
 [5,]  71  5
 [6,]  50  6
 [7,]  49  7
 [8,]  31  8
 [9,]  28  9
[10,]  17 10
[11,]  16 11
[12,]  11 12
```

```
plot(x,y, pch=16, col=2)
```




```
cor.test(x,y)
yt<-log(y)
plot(x,yt, pch=16, col=2)
cor.test(x,yt)
```

```
        Pearson's product-moment correlation

data:  x and yt
t = -23.4591, df = 10, p-value = 4.489e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9975654 -0.9672859
sample estimates:
       cor
-0.9910365
```

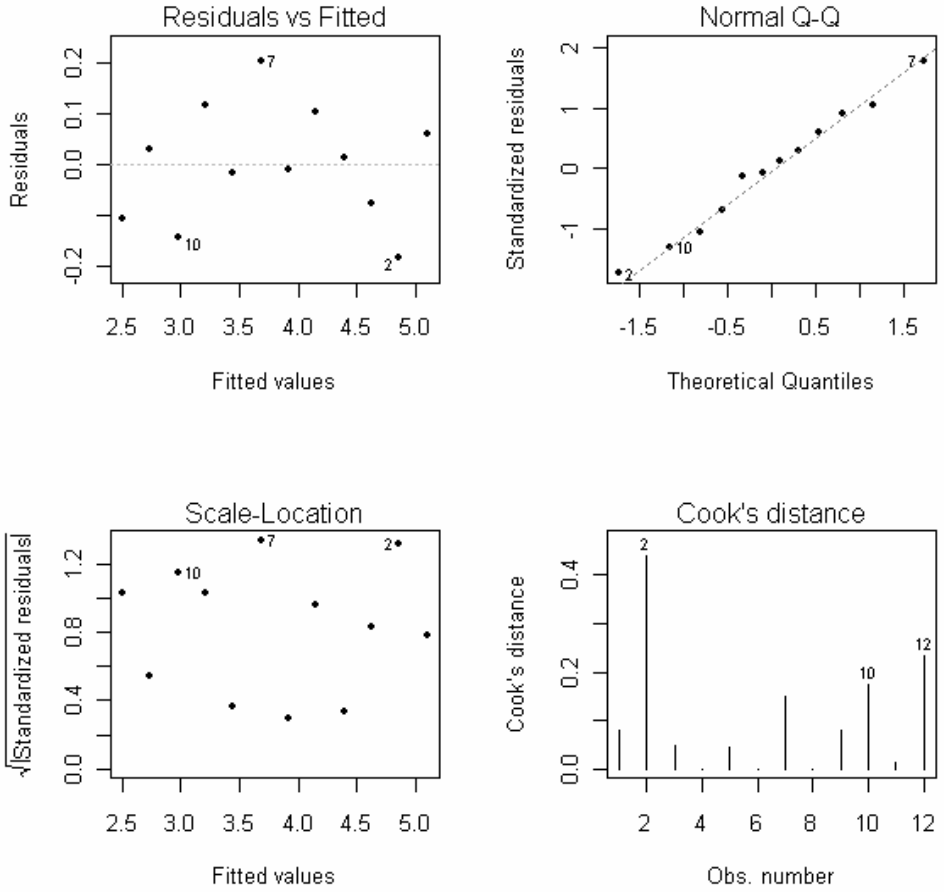



```
mod1<-lm(yt~x)
summary(mod1)

Call:
lm(formula = yt ~ x)

Residuals:
      Min        1Q    Median        3Q       Max
-0.184303 -0.083994  0.001453  0.072825  0.206246

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.33878    0.07409   72.05 6.47e-15 ***
```

```
x           -0.23617    0.01007  -23.46 4.49e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1204 on 10 degrees of freedom
Multiple R-squared: 0.9822,     Adjusted R-squared: 0.9804
F-statistic: 550.3 on 1 and 10 DF,  p-value: 4.489e-10
```

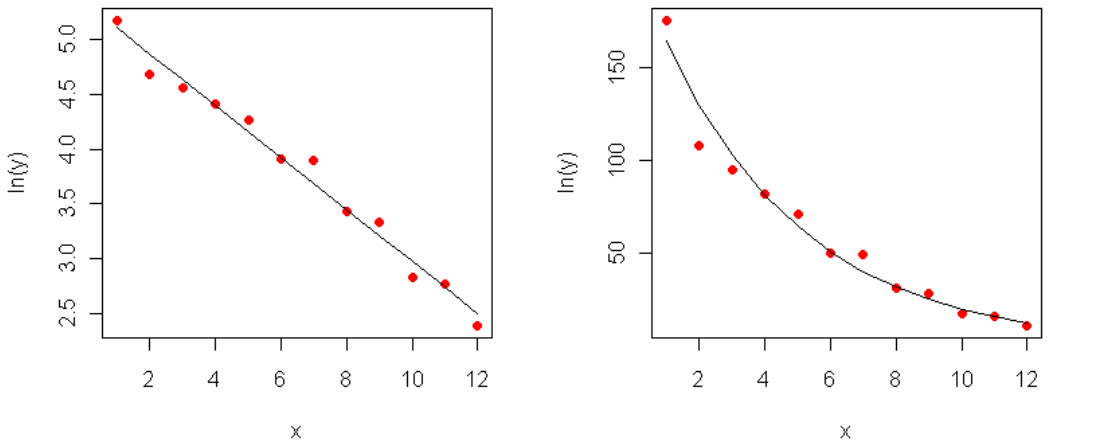
```
 anova(mod1)
Analysis of Variance Table
Response: yt
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 7.9761  7.9761  550.33 4.489e-10 ***
Residuals 10 0.1449  0.0145
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
 par(mfrow=c(2,2))
 plot(mod1, which=c(1:4), add.smooth=FALSE, pch=20)
 shapiro.test(mod1$resid)

        Shapiro-Wilk normality test

data:  mod1$resid
W = 0.9852, p-value = 0.9968
```

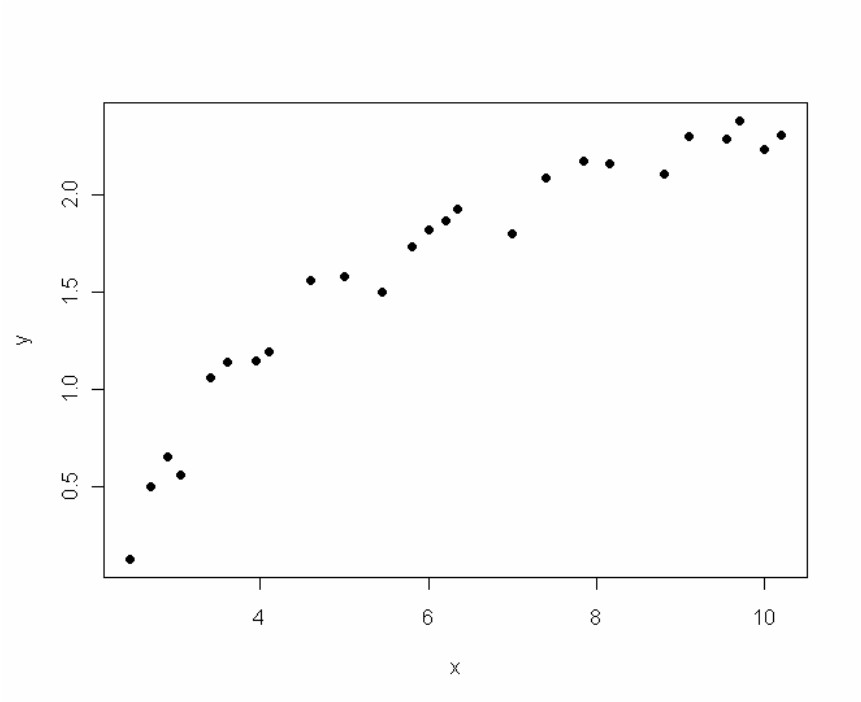



```
par(mfrow=c(1,2))
xe<-min(x):max(x)
yte<-mod1$coef[1] + mod1$coef[2]*xe
plot(x, yt, pch=16, col=4, ylab="ln(y)")
lines(xe, yte)
```

```
plot(x, y, pch=16, col=4, ylab="ln(y)")
ye<-exp(mod1$coef[1])*exp(mod1$coef[2]*xe)
lines(xe, ye)
```




```
#####################################################################
Exemplo 7 – Transformação em X
Y = energia elétrica gerada por um moinho de vento
X = velocidade do vento (em MPH)
Fonte: Montgomery e Peck, 1992.
#####################################################################

 rm(list = ls())
 ex7<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo7s.txt",h=T)
 attach(ex7)
 plot(x,y, pch=16)
```




```
## transformacao em x: procedimento de Box-Tidwell
## chute inicial: alpha0 = 1

mod1<-lm(y~x)
summary(mod1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59869 -0.14099  0.06059  0.17262  0.32184

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.13088    0.12599   1.039     0.31
x            0.24115    0.01905  12.659 7.55e-12 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2361 on 23 degrees of freedom
Multiple R-squared: 0.8745,     Adjusted R-squared: 0.869
F-statistic: 160.3 on 1 and 23 DF,  p-value: 7.546e-12


 w<-x*log(x)
 mod2<-lm(y~x+w)
 summary(mod2)

Residuals:
      Min        1Q    Median        3Q       Max
-0.223004 -0.029442  0.009955  0.048323  0.181553

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.41684    0.28512  -8.477 2.23e-08 ***
x            1.53443    0.14189  10.814 2.85e-10 ***
w           -0.46260    0.05065  -9.132 6.13e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1103 on 22 degrees of freedom
Multiple R-squared: 0.9738,     Adjusted R-squared: 0.9714
F-statistic: 408.9 on 2 and 22 DF,  p-value: < 2.2e-16

 alpha1<-(mod2$coef[3]/mod1$coef[2]) + 1
 alpha1
 -0.918302

 x1<-x^alpha1
 mod11<-lm(y~x1)
 summary(mod11)

Residuals:
     Min       1Q   Median       3Q      Max
-0.19539 -0.06179  0.01245  0.07813  0.12392
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.10386    0.04748   65.38   <2e-16 ***
x1          -6.67842    0.19538  -34.18   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09258 on 23 degrees of freedom
Multiple R-squared: 0.9807,     Adjusted R-squared: 0.9799
F-statistic:  1168 on 1 and 23 DF,  p-value: < 2.2e-16
```

```
w1<-x1*log(x1)
mod22<-lm(y~x1+w1)
summary(mod22)
```

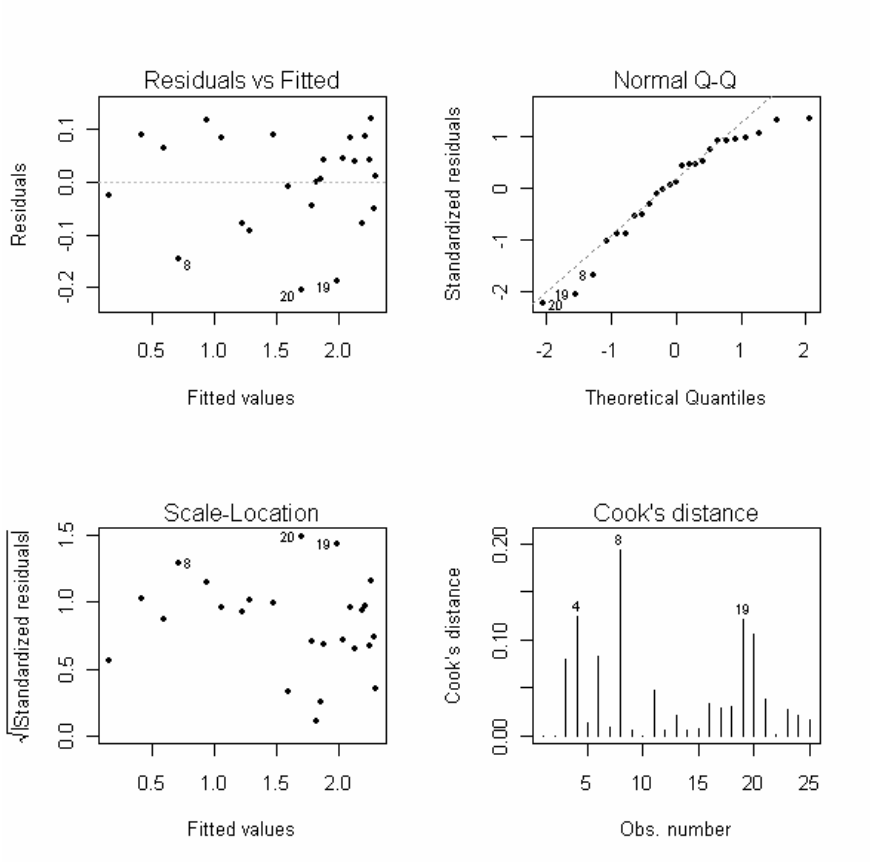
```
Call:
lm(formula = y ~ x1 + w1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.18507 -0.07049  0.01960  0.06727  0.12975

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.2409     0.2708  11.968 4.18e-11 ***
x1           -6.4445     0.4962 -12.987 8.58e-12 ***
w1            0.5994     1.1652   0.514    0.612
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0941 on 22 degrees of freedom
Multiple R-squared: 0.9809,     Adjusted R-squared: 0.9792
F-statistic: 565.6 on 2 and 22 DF,  p-value: < 2.2e-16
```

```
alpha2<-(mod22$coef[3]/mod11$coef[2]) + alpha1
alpha2
```

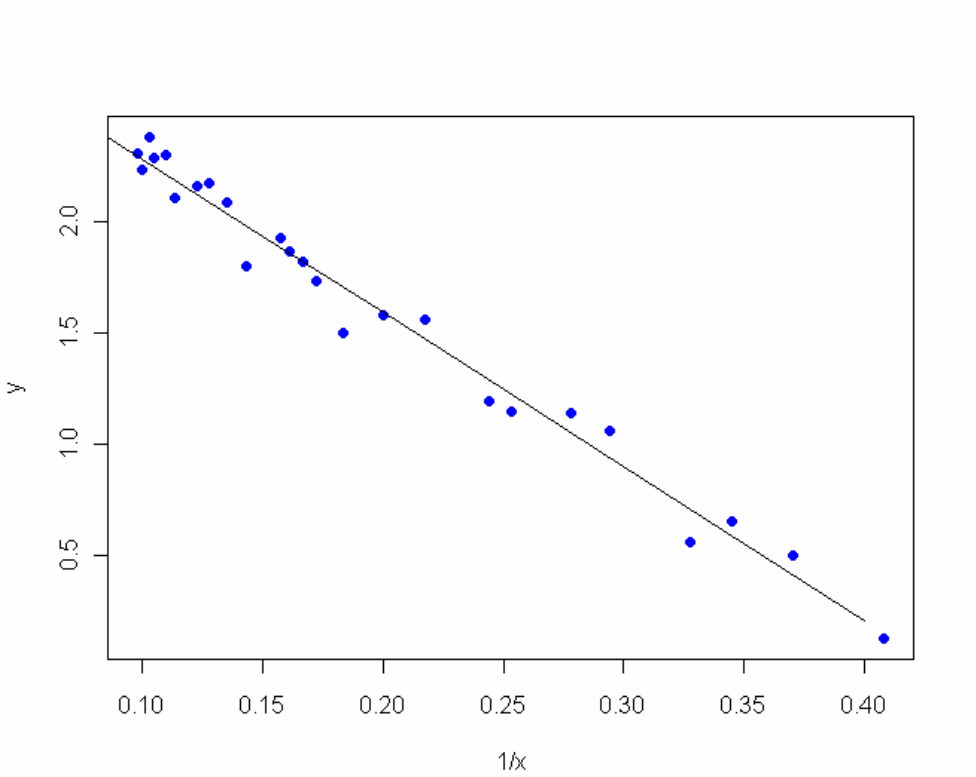
```
-1.008055
```

 # transformacao indicada: $x^{-1} = 1/x$

```
xt<-1/x
plot(xt,y, pch=16,col=2)
mod3<-lm(y~xt)
summary(mod3)

Residuals:
     Min       1Q   Median       3Q      Max
-0.20547 -0.04941  0.01100  0.08352  0.12204

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.9789     0.0449   66.34   <2e-16 ***
xt           -6.9345     0.2064  -33.59   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09417 on 23 degrees of freedom
Multiple R-squared:  0.98,      Adjusted R-squared: 0.9792
F-statistic:  1128 on 1 and 23 DF,  p-value: < 2.2e-16

par(mfrow=c(2,2))
plot(mod3, which=c(1:4), add.smooth=FALSE, pch=20)
shapiro.test(mod3$resid)

        Shapiro-Wilk normality test
data:  mod3$resid
W = 0.9292, p-value = 0.0835
```

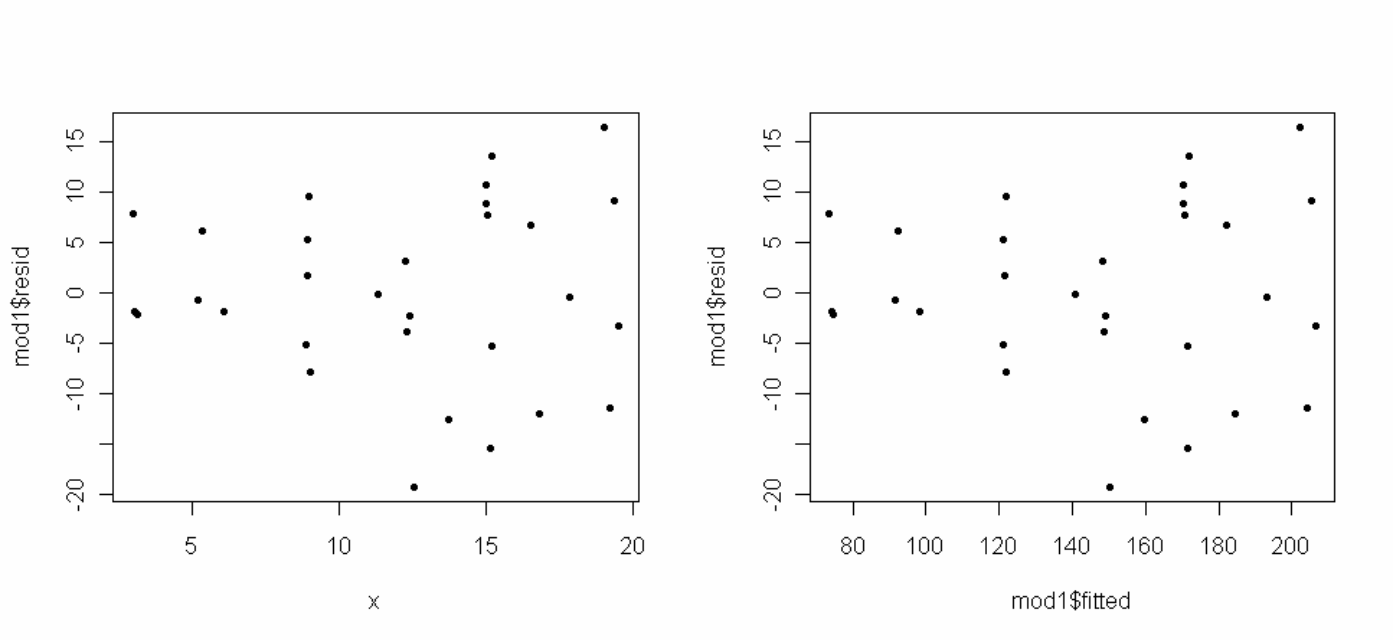



```
par(mfrow=c(1,1))
plot(xt,y, pch=16, col=4, xlab="1/x")
xe<-seq(0,0.41,by=0.05)
ye<-mod3$coef[1]+mod3$coef[2]*xe
lines(xe,ye)
```



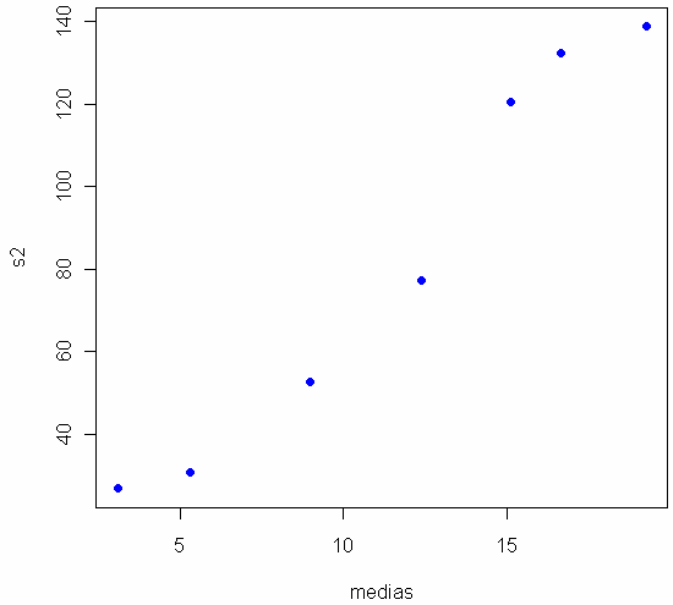
Exemplo 4: Mínimos Quadrados Ponderados

```
rm(list = ls())
ex4<-read.table("http://www.ufpr.br/~giolo/CE071/Exemplos/Exemplo4s.txt",h=T)
attach(ex4)
mod1<-lm(y~x)
par(mfrow=c(1,2))
plot(x, mod1$resid, pch = 20)
plot(mod1$fitted, mod1$resid, pch = 20)
```




```
m1<-mean(x[1:3])
s1<-var(y[1:3])
m2<-mean(x[4:5])
s2<-var(y[4:5])
m3<-mean(x[7:11])
s3<-var(y[7:11])
m4<-mean(x[13:16])
s4<-var(y[13:16])
m5<-mean(x[18:23])
s5<-var(y[18:23])
m6<-mean(x[24:25])
s6<-var(y[24:25])
m7<-mean(x[27:30])
s7<-var(y[27:30])
medias<-as.vector(c(m1,m2,m3,m4,m5,m6,m7))
```

```
 s2<-as.vector(c(s1,s2,s3,s4,s5,s6,s7))
 medias
[1]  3.078333  5.287500  8.955000 12.377500 15.095000 16.650000 19.262500
 s2
[1]  26.79462  30.77201  52.80369  77.28016 120.57106 132.38899 138.85687

 plot(medias, s2, pch=16, col=4)
```

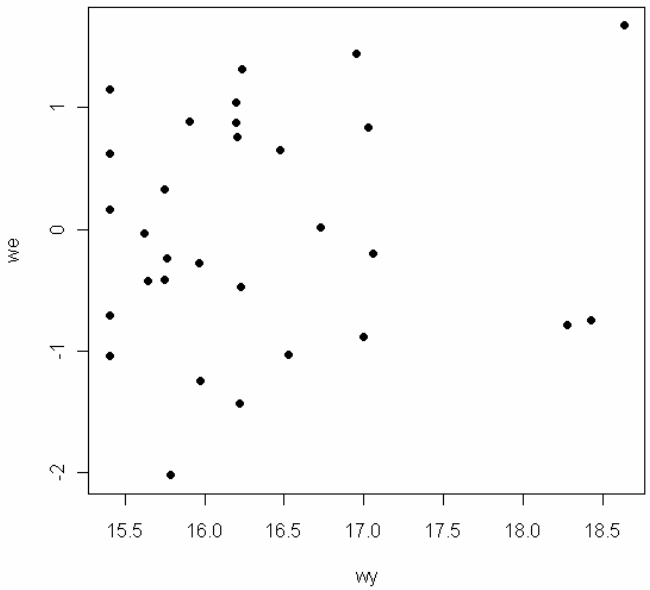



```
 lm(s2~medias)

 Coefficients:
(Intercept)       medias
     -7.376        7.820

 aux<--7.376+7.820*x
 wi<-1/aux
```

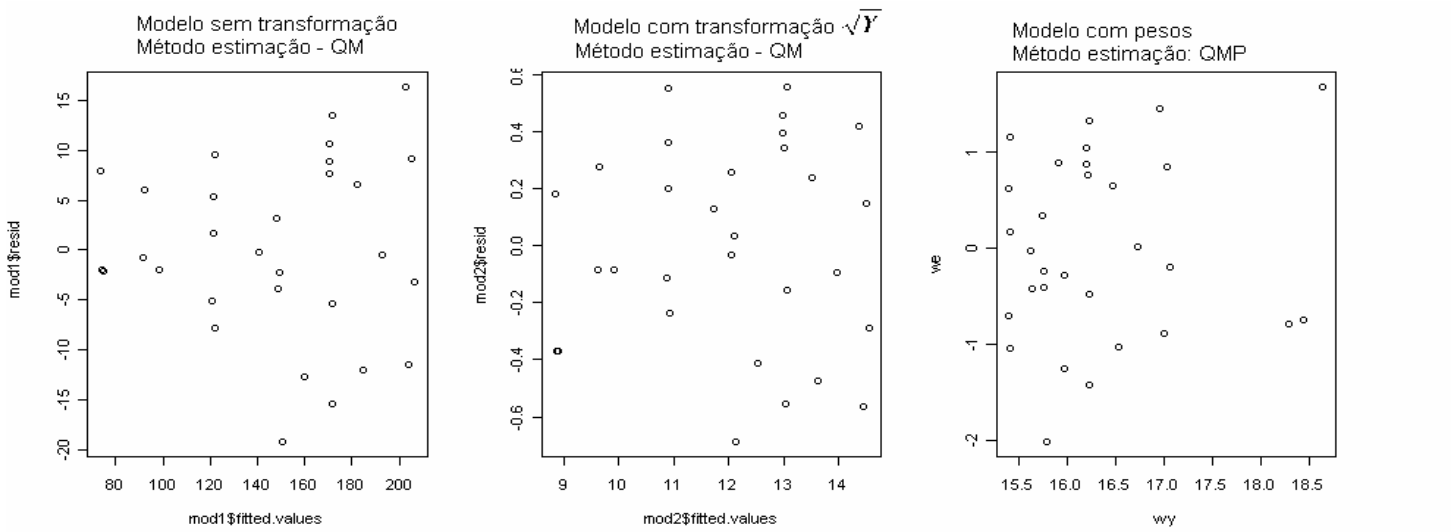
```
 wi
 [1] 0.062173589 0.057947500 0.059706126 0.029865456 0.029018311 0.024846079
 [7] 0.016021148 0.015842513 0.016101841 0.015971124 0.015872008 0.012293787
[13] 0.011284835 0.011161711 0.011041245 0.011250087 0.010024259 0.009097194
[19] 0.008985331 0.009100431 0.009064950 0.008969575 0.009001143 0.008064516
[25] 0.008220034 0.007572625 0.006891134 0.007004371 0.007081952 0.006947291
```

```
 mod3<-lm(y~x,weights=wi)
 summary(mod3)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0214 -0.7388 -0.1134  0.8181  1.6763

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.9746     2.5073   20.33   <2e-16 ***
x             7.9222     0.2532   31.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9619 on 28 degrees of freedom
Multiple R-squared: 0.9722,     Adjusted R-squared: 0.9712
F-statistic: 978.6 on 1 and 28 DF,  p-value: < 2.2e-16

 anova(mod3)

Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 905.50  905.50  978.63 < 2.2e-16 ***
Residuals 28  25.91    0.93
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 rqwi<-sqrt(wi)
 we<-rqwi*mod3$resid
 wx<-rqwi*x
```

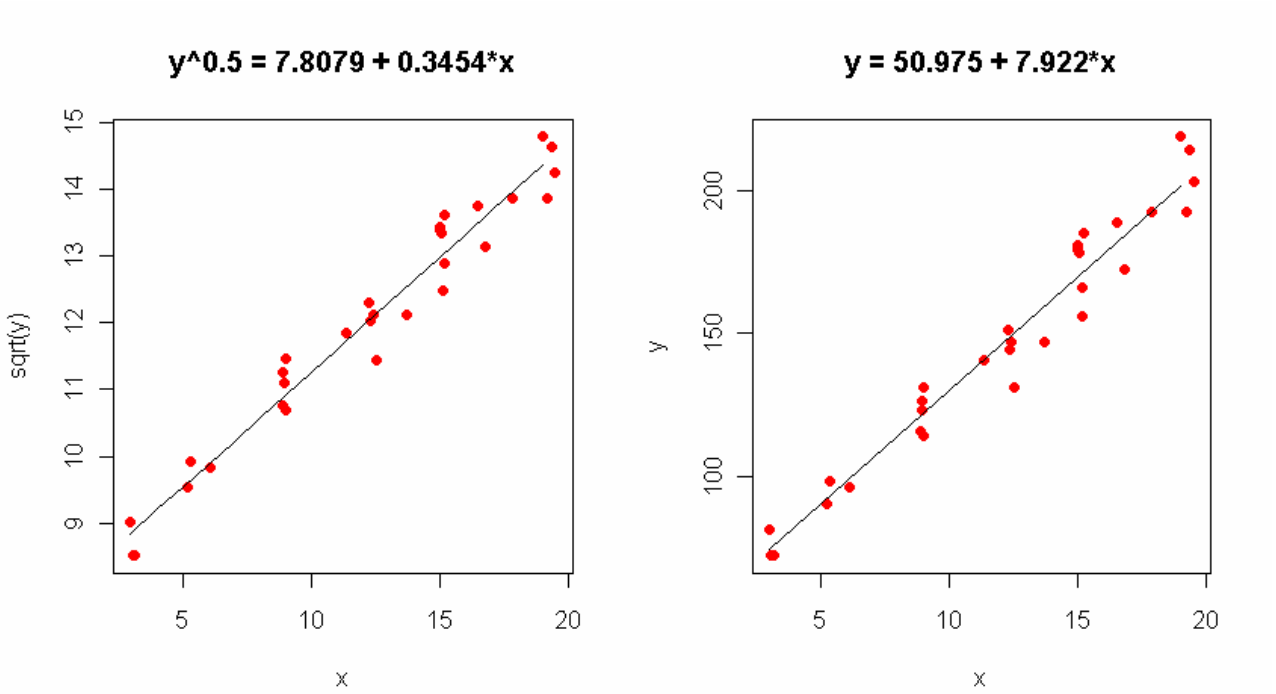
```
wy<-rqwi*mod3$fitted.values
plot(wy,we, pch=16)
```




```
mod1<-lm(y~x)
mod2<-lm(sqrt(y)~x)
mod3<-lm(y~x,weights=wi)
par(mfrow=c(1,3))
plot(mod1$fitted.values,mod1$resid)
plot(mod2$fitted.values,mod2$resid)
plot(wy,we)
```




```
par(mfrow=c(1,2))
plot(x,sqrt(y), pch=16,col=2)
xe<- min(x):max(x)
ye2<- mod2$coef[1]+mod2$coef[2]*xe
ye3<- mod3$coef[1]+mod3$coef[2]*xe
lines(xe,ye2)
title("y^0.5 = 7.8079 + 0.3454*x")
plot(x,y, pch=16,col=2)
lines(xe,ye3)
title("y = 50.975 + 7.922*x")
```


y^0.5 = 7.8079 + 0.3454*x
sqrt(y)
x
y = 50.975 + 7.922*x
y
x