Exercicio3.R

Jenifer Sofia Ovejero (22200500)

2023-10-18

a) Utilizando o método de seleção de variáveis “forward”, ajuste o modelo mais adequado para o conjunto dos dados.

```
library(data.table)
base <- fread(input = paste0("selecao.csv"), header = T, na.strings = "NA",
data.table = FALSE, dec=",")
names(base)
```

```
##  [1] "x1" "x2" "x3" "x4" "x5" "x6" "x7" "x8" "x9" "y"
```

```
# seleção de variaveis modelo 1
m0 = lm(y ~ 1, data = base)
m1 = step(m0,list(lower = ~ 1,
                  upper = ~ x1+x2+x3+x4+x5+x6+x7+x8+x9),
          direction = "forward")
```

```
## Start:  AIC=5655.48
## y ~ 1
##
##        Df Sum of Sq   RSS    AIC
## + x7    1     53916 21387 3863.8
## + x9    1     53030 22274 3921.7
## + x8    1     42439 32865 4476.0
## + x5    1     33450 41854 4820.5
## + x6    1     33024 42279 4834.9
## + x4    1     17408 57896 5282.9
## + x2    1       257 75047 5652.6
## + x3    1       253 75050 5652.7
## <none>              75304 5655.5
## + x1    1        32 75272 5656.9
##
## Step:  AIC=3863.78
## y ~ x7
##
##        Df Sum of Sq   RSS    AIC
## + x9    1   1283.44 20104 3777.6
## + x8    1   1260.79 20126 3779.2
## + x6    1    783.20 20604 3812.6
## + x5    1    420.78 20966 3837.5
## + x1    1     89.07 21298 3859.8
## + x4    1     53.18 21334 3862.2
## <none>              21387 3863.8
```

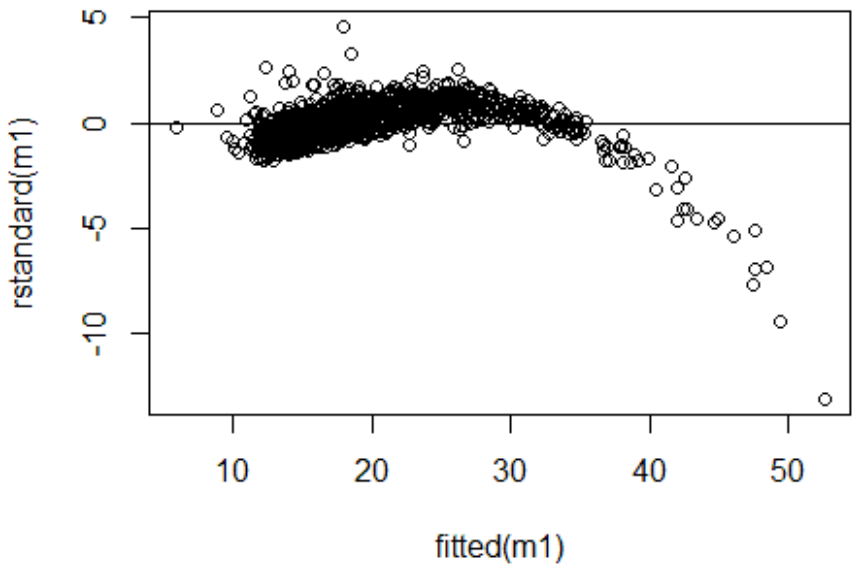
```
## + x3    1      8.26 21379 3865.2
## + x2    1      3.51 21384 3865.6
##
## Step:  AIC=3777.6
## y ~ x7 + x9
##
##        Df Sum of Sq   RSS    AIC
## + x6    1   123.072 19981 3770.8
## + x4    1    81.652 20022 3773.8
## + x1    1    54.956 20049 3775.7
## + x8    1    36.266 20068 3777.0
## <none>              20104 3777.6
## + x5    1     4.217 20100 3779.3
## + x2    1     0.713 20103 3779.5
## + x3    1     0.536 20103 3779.6
##
## Step:  AIC=3770.85
## y ~ x7 + x9 + x6
##
##        Df Sum of Sq   RSS    AIC
## + x4    1   1080.45 18900 3693.6
## + x1    1     81.08 19900 3767.1
## + x8    1     36.90 19944 3770.2
## + x5    1     34.74 19946 3770.4
## <none>              19981 3770.8
## + x2    1      0.78 19980 3772.8
## + x3    1      0.23 19981 3772.8
##
## Step:  AIC=3693.63
## y ~ x7 + x9 + x6 + x4
##
##        Df Sum of Sq   RSS    AIC
## + x5    1   236.064 18664 3677.7
## + x1    1    62.304 18838 3690.9
## + x8    1    36.381 18864 3692.9
## <none>              18900 3693.6
## + x3    1     1.815 18898 3695.5
## + x2    1     1.213 18899 3695.5
##
## Step:  AIC=3677.72
## y ~ x7 + x9 + x6 + x4 + x5
##
##        Df Sum of Sq   RSS    AIC
## + x1    1    52.902 18611 3675.7
## + x8    1    40.046 18624 3676.7
## <none>              18664 3677.7
## + x3    1     6.807 18657 3679.2
## + x2    1     1.911 18662 3679.6
##
## Step:  AIC=3675.67
```

```
## y ~ x7 + x9 + x6 + x4 + x5 + x1
##
##        Df Sum of Sq   RSS    AIC
## + x8    1    41.362 18570 3674.5
## <none>              18611 3675.7
## + x3    1     7.054 18604 3677.1
## + x2    1     3.030 18608 3677.4
##
## Step:  AIC=3674.5
## y ~ x7 + x9 + x6 + x4 + x5 + x1 + x8
##
##        Df Sum of Sq   RSS    AIC
## <none>              18570 3674.5
## + x3    1    6.5176 18563 3676.0
## + x2    1    3.3522 18567 3676.2
```

Modelo = y ~ x7 + x9 + x6 + x4 + x5 + x1 + x8

b) Faça a análise de resíduos para checar se o modelo está bem ajustado, caso o modelo não esteja bem ajustado, identifique o que é possível fazer para que o modelo fique bem ajustado.

```
# Qualidade do Ajuste (análise de resíduos)
plot(fitted(m1), rstandard(m1))
abline(0, 0)
```




```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: y
##             Df Sum Sq Mean Sq   F value    Pr(>F)
## x7           1  53916   53916 4114.1621 < 2.2e-16 ***
## x9           1   1283    1283   97.9346 < 2.2e-16 ***
## x6           1    123     123    9.3911  0.002222 **
## x4           1   1080    1080   82.4448 < 2.2e-16 ***
## x5           1    236     236   18.0132 2.336e-05 ***
## x1           1     53      53    4.0368  0.044708 *
## x8           1     41      41    3.1562  0.075854 .
## Residuals 1417  18570      13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = y ~ x7 + x9 + x6 + x4 + x5 + x1 + x8, data = base)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.172  -1.925   0.169   2.063  16.032
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.5374685  1.4415446  -4.535 6.24e-06 ***
## x7          -0.2638935  0.4659336  -0.566   0.5712
## x9           0.9462288  0.4643885   2.038   0.0418 *
## x6           0.2664206  0.0278735   9.558  < 2e-16 ***
## x4          -0.2068910  0.0211328  -9.790  < 2e-16 ***
## x5           0.3019813  0.0720512   4.191 2.95e-05 ***
## x1          -0.0011609  0.0005707  -2.034   0.0421 *
## x8          -0.8250781  0.4644232  -1.777   0.0759 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.62 on 1417 degrees of freedom
## Multiple R-squared:  0.7534, Adjusted R-squared:  0.7522
## F-statistic: 618.4 on 7 and 1417 DF,  p-value: < 2.2e-16
```

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: y
##             Df Sum Sq Mean Sq   F value    Pr(>F)
## x7           1  53916   53916 4114.1621 < 2.2e-16 ***
## x9           1   1283    1283   97.9346 < 2.2e-16 ***
## x6           1    123     123    9.3911  0.002222 **
```

```
## x4           1   1080    1080   82.4448 < 2.2e-16 ***
## x5           1    236     236   18.0132 2.336e-05 ***
## x1           1     53      53    4.0368  0.044708 *
## x8           1     41      41    3.1562  0.075854 .
## Residuals 1417  18570      13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

O modelo não está bem ajustado, pelo gráfico é possível notar que os pontos não se encontram entre 3 e -3 e não estão bem distribuídos. Para o ajuste do modelo as variáveis serão transformadas para seu quadrado.

```
# transformação das variáveis
base$x1sqrt = base$x1 ^ 2
base$x2sqrt = base$x2 ^ 2
base$x3sqrt = base$x3 ^ 2
base$x4sqrt = base$x4 ^ 2
base$x5sqrt = base$x5 ^ 2
base$x6sqrt = base$x6 ^ 2
base$x7sqrt = base$x7 ^ 2
base$x8sqrt = base$x8 ^ 2
base$x9sqrt = base$x9 ^ 2
```

c) Se necessário fazer alguma transformação ou acrescentar algum termo no modelo, utilize novamente o método de seleção de variáveis “forward” e refaça a análise do item (b).

```
# seleção de variaveis modelo 2
m0 = lm(y ~ 1, data = base)
m2 = step(m0, list(lower = ~ 1,
                   upper = ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +
                     x1sqrt + x2sqrt + x3sqrt + x4sqrt + x5sqrt + x6sqrt +
x7sqrt + x8sqrt + x9sqrt),
          direction = "forward")
```

```
## Start:  AIC=5655.48
## y ~ 1
##
##          Df Sum of Sq   RSS    AIC
## + x7      1     53916 21387 3863.8
## + x9      1     53030 22274 3921.7
## + x8      1     42439 32865 4476.0
## + x7sqrt  1     36918 38385 4697.2
## + x5      1     33450 41854 4820.5
## + x6      1     33024 42279 4834.9
## + x6sqrt  1     31826 43477 4874.8
## + x9sqrt  1     30948 44356 4903.2
## + x5sqrt  1     29295 46008 4955.4
## + x8sqrt  1     21855 53449 5169.0
## + x4      1     17408 57896 5282.9
## + x4sqrt  1     14964 60339 5341.8
```

```
## + x2      1       257 75047 5652.6
## + x3      1       253 75050 5652.7
## + x3sqrt  1       251 75052 5652.7
## + x2sqrt  1       193 75110 5653.8
## <none>                75304 5655.5
## + x1      1        32 75272 5656.9
## + x1sqrt  1         4 75299 5657.4
##
## Step:  AIC=3863.78
## y ~ x7
##
##          Df Sum of Sq     RSS    AIC
## + x7sqrt  1   12083.5  9303.7 2679.6
## + x9sqrt  1    2311.6 19075.6 3702.8
## + x9      1    1283.4 20103.8 3777.6
## + x8      1    1260.8 20126.4 3779.2
## + x6      1     783.2 20604.0 3812.6
## + x8sqrt  1     487.6 20899.6 3832.9
## + x6sqrt  1     445.6 20941.6 3835.8
## + x5      1     420.8 20966.4 3837.5
## + x1      1      89.1 21298.1 3859.8
## + x4      1      53.2 21334.0 3862.2
## + x1sqrt  1      49.4 21337.8 3862.5
## + x5sqrt  1      43.4 21343.8 3862.9
## <none>                21387.2 3863.8
## + x3      1       8.3 21378.9 3865.2
## + x4sqrt  1       8.0 21379.2 3865.3
## + x3sqrt  1       3.8 21383.4 3865.5
## + x2      1       3.5 21383.7 3865.6
## + x2sqrt  1       1.6 21385.6 3865.7
##
## Step:  AIC=2679.65
## y ~ x7 + x7sqrt
##
##          Df Sum of Sq    RSS    AIC
## + x9      1    3712.2 5591.5 1956.1
## + x8      1    3693.3 5610.4 1960.9
## + x9sqrt  1    1150.4 8153.3 2493.6
## + x8sqrt  1    1100.6 8203.1 2502.2
## + x5      1     632.8 8670.9 2581.3
## + x6      1     615.1 8688.7 2584.2
## + x6sqrt  1     600.9 8702.8 2586.5
## + x5sqrt  1     486.0 8817.7 2605.2
## + x4      1     170.2 9133.6 2655.3
## + x4sqrt  1     132.4 9171.3 2661.2
## + x1      1     116.0 9187.7 2663.8
## + x1sqrt  1      95.9 9207.8 2666.9
## + x2      1      42.4 9261.4 2675.1
## + x2sqrt  1      31.0 9272.7 2676.9
## <none>                9303.7 2679.7
```

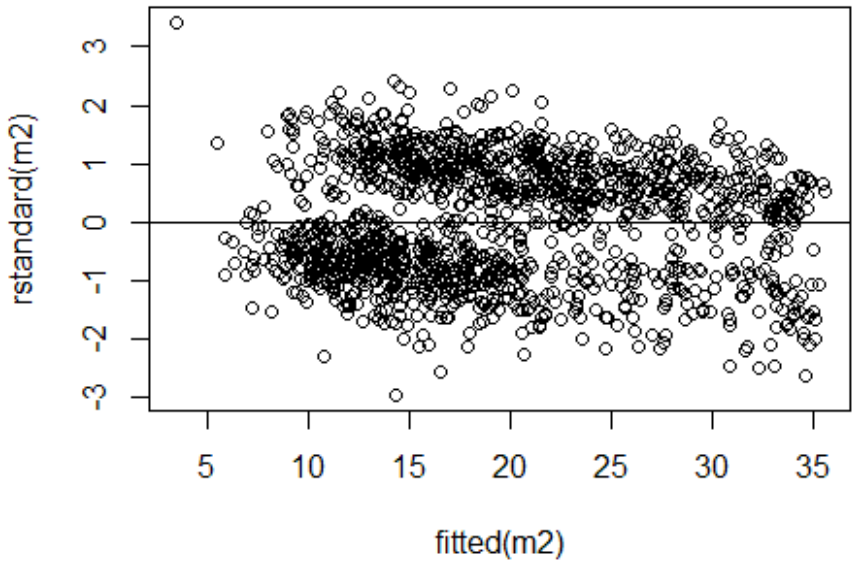
```
## + x3      1      11.7 9292.0 2679.8
## + x3sqrt  1       9.0 9294.7 2680.3
##
## Step:  AIC=1956.1
## y ~ x7 + x7sqrt + x9
##
##          Df Sum of Sq    RSS     AIC
## + x9sqrt  1    3533.6 2057.9  533.73
## + x8sqrt  1    3248.6 2343.0  718.57
## + x4sqrt  1     355.4 5236.2 1864.53
## + x4      1     213.6 5378.0 1902.60
## + x5sqrt  1     161.1 5430.4 1916.43
## + x6sqrt  1     160.6 5430.9 1916.56
## + x6      1      85.1 5506.5 1936.25
## + x5      1      67.6 5523.9 1940.76
## + x1      1      54.7 5536.8 1944.08
## + x1sqrt  1      40.4 5551.2 1947.77
## + x2      1      27.9 5563.7 1950.97
## + x2sqrt  1      20.9 5570.7 1952.77
## <none>                5591.5 1956.10
## + x8      1       3.3 5588.2 1957.25
## + x3sqrt  1       0.2 5591.4 1958.05
## + x3      1       0.1 5591.5 1958.08
##
## Step:  AIC=533.73
## y ~ x7 + x7sqrt + x9 + x9sqrt
##
##          Df Sum of Sq    RSS    AIC
## + x4      1    610.76 1447.2  34.00
## + x4sqrt  1    505.00 1552.9 134.51
## + x6      1    429.49 1628.4 202.17
## + x6sqrt  1    406.51 1651.4 222.13
## + x5      1    287.42 1770.5 321.36
## + x5sqrt  1    219.01 1838.9 375.39
## + x1      1     52.55 2005.4 498.87
## + x1sqrt  1     40.86 2017.1 507.15
## + x2      1     18.68 2039.2 522.73
## + x2sqrt  1     14.59 2043.3 525.59
## + x3      1      6.34 2051.6 531.33
## + x3sqrt  1      6.21 2051.7 531.42
## <none>                2057.9 533.73
## + x8      1      2.02 2055.9 534.33
## + x8sqrt  1      0.17 2057.8 535.61
##
## Step:  AIC=34
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4
##
##          Df Sum of Sq    RSS    AIC
## + x4sqrt  1   25.6879 1421.5 10.474
## + x2      1   18.6014 1428.6 17.561
```

```
## + x5sqrt  1   16.0948 1431.1 20.059
## + x1      1   15.7760 1431.4 20.376
## + x2sqrt  1   15.5009 1431.7 20.650
## + x1sqrt  1   14.1643 1433.0 21.980
## + x5      1    5.3505 1441.8 30.718
## + x8sqrt  1    4.1251 1443.0 31.928
## + x8      1    3.2531 1443.9 32.789
## <none>                1447.2 33.996
## + x6sqrt  1    0.6553 1446.5 35.350
## + x3sqrt  1    0.2214 1447.0 35.778
## + x3      1    0.1260 1447.0 35.872
## + x6      1    0.0988 1447.1 35.899
##
## Step:  AIC=10.47
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt
##
##          Df Sum of Sq    RSS     AIC
## + x2      1   16.8783 1404.6 -4.5471
## + x2sqrt  1   13.9478 1407.5 -1.5771
## + x1      1   10.9125 1410.6  1.4925
## + x1sqrt  1   10.7150 1410.8  1.6920
## + x5sqrt  1    6.8710 1414.6  5.5695
## + x8sqrt  1    6.4964 1415.0  5.9468
## + x8      1    3.5362 1417.9  8.9249
## + x5      1    3.1005 1418.4  9.3627
## <none>                1421.5 10.4743
## + x6      1    0.4193 1421.1 12.0539
## + x6sqrt  1    0.3498 1421.1 12.1236
## + x3sqrt  1    0.1937 1421.3 12.2801
## + x3      1    0.1701 1421.3 12.3038
##
## Step:  AIC=-4.55
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2
##
##          Df Sum of Sq    RSS     AIC
## + x5sqrt  1    7.0259 1397.6 -9.6929
## + x8sqrt  1    6.1363 1398.5 -8.7862
## + x8      1    3.6481 1401.0 -6.2530
## + x5      1    3.4024 1401.2 -6.0031
## + x1      1    3.2011 1401.4 -5.7984
## + x1sqrt  1    3.1173 1401.5 -5.7132
## <none>                1404.6 -4.5471
## + x2sqrt  1    1.0229 1403.6 -3.5852
## + x3sqrt  1    0.3934 1404.2 -2.9462
## + x6      1    0.3806 1404.2 -2.9332
## + x3      1    0.3631 1404.2 -2.9155
## + x6sqrt  1    0.3112 1404.3 -2.8629
##
## Step:  AIC=-9.69
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 + x5sqrt
```

```
##
##          Df Sum of Sq    RSS      AIC
## + x5      1    9.5645 1388.0 -17.4787
## + x8sqrt  1    5.0152 1392.6 -12.8158
## + x8      1    3.4570 1394.1 -11.2221
## + x1sqrt  1    2.9101 1394.7 -10.6633
## + x1      1    2.8592 1394.7 -10.6112
## <none>                1397.6  -9.6929
## + x2sqrt  1    0.8220 1396.8  -8.5313
## + x6      1    0.7443 1396.8  -8.4521
## + x3      1    0.6349 1396.9  -8.3405
## + x3sqrt  1    0.6094 1397.0  -8.3144
## + x6sqrt  1    0.5502 1397.0  -8.2541
##
## Step:  AIC=-17.48
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 + x5sqrt + x5
##
##          Df Sum of Sq    RSS     AIC
## + x8sqrt  1    5.7644 1382.2 -21.409
## + x8      1    3.7916 1384.2 -19.377
## + x1sqrt  1    3.0453 1385.0 -18.609
## + x1      1    2.9664 1385.0 -18.527
## <none>                1388.0 -17.479
## + x6      1    1.6185 1386.4 -17.141
## + x6sqrt  1    1.1214 1386.9 -16.630
## + x2sqrt  1    0.7897 1387.2 -16.290
## + x3      1    0.3904 1387.6 -15.880
## + x3sqrt  1    0.3625 1387.7 -15.851
##
## Step:  AIC=-21.41
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 + x5sqrt + x5 +
##     x8sqrt
##
##          Df Sum of Sq    RSS     AIC
## + x8      1    4.8636 1377.4 -24.432
## + x1sqrt  1    2.7895 1379.5 -22.288
## + x1      1    2.6732 1379.6 -22.168
## <none>                1382.2 -21.409
## + x6      1    1.1017 1381.1 -20.545
## + x6sqrt  1    0.7646 1381.5 -20.198
## + x2sqrt  1    0.6816 1381.6 -20.112
## + x3      1    0.4716 1381.8 -19.895
## + x3sqrt  1    0.4469 1381.8 -19.870
##
## Step:  AIC=-24.43
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 + x5sqrt + x5 +
##     x8sqrt + x8
##
##          Df Sum of Sq    RSS     AIC
## + x1sqrt  1   2.61471 1374.8 -25.140
```

```
## + x1      1   2.48650 1374.9 -25.007
## <none>                1377.4 -24.432
## + x6      1   1.00737 1376.4 -23.474
## + x2sqrt  1   0.72949 1376.7 -23.187
## + x6sqrt  1   0.68601 1376.7 -23.142
## + x3      1   0.52423 1376.9 -22.974
## + x3sqrt  1   0.49806 1376.9 -22.947
##
## Step:  AIC=-25.14
## y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 + x5sqrt + x5 +
##     x8sqrt + x8 + x1sqrt
##
##          Df Sum of Sq    RSS     AIC
## <none>                1374.8 -25.140
## + x6      1   1.10736 1373.7 -24.288
## + x6sqrt  1   0.76060 1374.0 -23.928
## + x3      1   0.50175 1374.3 -23.660
## + x2sqrt  1   0.45557 1374.3 -23.612
## + x3sqrt  1   0.44316 1374.3 -23.599
## + x1      1   0.08695 1374.7 -23.230
```

```r
# Qualidade do Ajuste (análise de resíduos)
plot(fitted(m2), rstandard(m2))
abline(0, 0)
```




```r
anova(m2)
```

```
## Analysis of Variance Table
##
## Response: y
##             Df Sum Sq Mean Sq    F value    Pr(>F)
## x7           1  53916   53916 55376.6783 < 2.2e-16 ***
## x7sqrt       1  12083   12083 12410.7327 < 2.2e-16 ***
## x9           1   3712    3712  3812.7111 < 2.2e-16 ***
## x9sqrt       1   3534    3534  3629.3205 < 2.2e-16 ***
## x4           1    611     611   627.3036 < 2.2e-16 ***
## x4sqrt       1     26      26    26.3836 3.190e-07 ***
## x2           1     17      17    17.3354 3.323e-05 ***
## x5sqrt       1      7       7     7.2162  0.007310 **
## x5           1     10      10     9.8235  0.001758 **
## x8sqrt       1      6       6     5.9205  0.015089 *
## x8           1      5       5     4.9953  0.025571 *
## x1sqrt       1      3       3     2.6855  0.101487
## Residuals 1412   1375       1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m2)
```

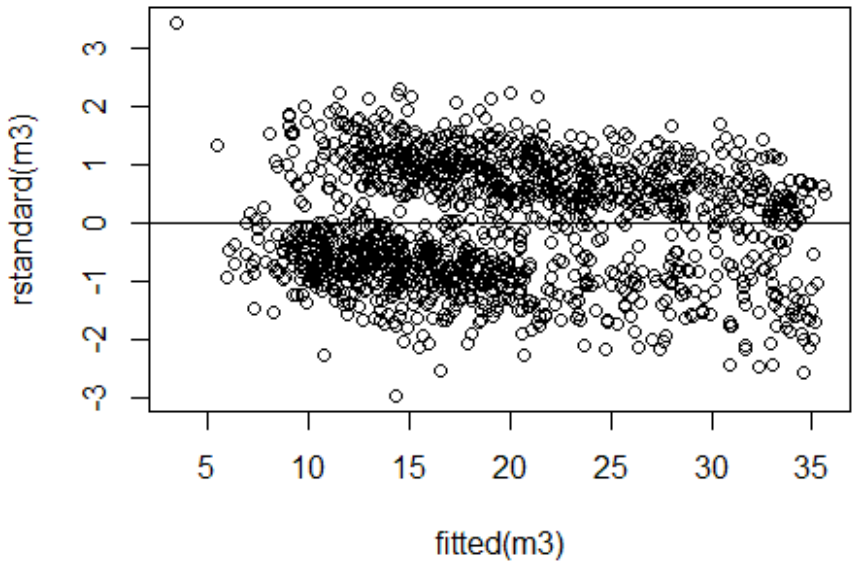
```
##
## Call:
## lm(formula = y ~ x7 + x7sqrt + x9 + x9sqrt + x4 + x4sqrt + x2 +
##     x5sqrt + x5 + x8sqrt + x8 + x1sqrt, data = base)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.88560 -0.80695  0.00462  0.83966  2.98555
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.942e-01  6.658e-01   0.292 0.770622
## x7           3.753e-01  1.312e-01   2.860 0.004293 **
## x7sqrt      -5.654e-03  1.510e-03  -3.743 0.000189 ***
## x9           1.144e+00  1.275e-01   8.967  < 2e-16 ***
## x9sqrt      -1.080e-02  6.363e-04 -16.974  < 2e-16 ***
## x4          -8.615e-02  1.356e-02  -6.355  2.8e-10 ***
## x4sqrt       1.726e-04  1.374e-04   1.256 0.209294
## x2          -1.883e-01  6.370e-02  -2.956 0.003163 **
## x5sqrt       7.326e-03  1.922e-03   3.811 0.000144 ***
## x5          -2.805e-01  8.386e-02  -3.344 0.000846 ***
## x8sqrt      -3.122e-03  1.204e-03  -2.593 0.009619 **
## x8           2.801e-01  1.276e-01   2.194 0.028363 *
## x1sqrt      -4.242e-07  2.588e-07  -1.639 0.101487
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9867 on 1412 degrees of freedom
```

```
## Multiple R-squared:  0.9817, Adjusted R-squared:  0.9816
## F-statistic:  6328 on 12 and 1412 DF,  p-value: < 2.2e-16
```

Note que agora o modelo está bem ajustado, entretanto existem variáveis que possuem p-valor superior a 0.05 (nível de significância), por esse motivo serão retiradas do modelo. Assim obtendo um modelo com menos variáveis e ajustado.

```
# Modelo Final
m3 = lm(y ~ x2 + x4 + x5 + x7 + x8 + x9 +  x5sqrt + x7sqrt + x8sqrt + x9sqrt,
data = base)

# Qualidade do Ajuste (análise de resíduos)
plot(fitted(m3), rstandard(m3))
abline(0, 0)
```




```
anova(m3)
```

```
## Analysis of Variance Table
##
## Response: y
##             Df  Sum Sq Mean Sq   F value    Pr(>F)
## x2           1   256.8   256.8   263.323 < 2.2e-16 ***
## x4           1 17256.2 17256.2 17691.116 < 2.2e-16 ***
## x5           1 18399.7 18399.7 18863.472 < 2.2e-16 ***
## x7           1 18671.3 18671.3 19141.865 < 2.2e-16 ***
## x8           1   858.6   858.6   880.262 < 2.2e-16 ***
## x9           1    61.4    61.4    62.988 4.201e-15 ***
```

```
## x5sqrt       1  5071.1  5071.1  5198.875 < 2.2e-16 ***
## x7sqrt       1  9911.7  9911.7 10161.556 < 2.2e-16 ***
## x8sqrt       1  3155.5  3155.5  3235.044 < 2.2e-16 ***
## x9sqrt       1   282.0   282.0   289.134 < 2.2e-16 ***
## Residuals 1414  1379.2     1.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m3)
```

```
##
## Call:
## lm(formula = y ~ x2 + x4 + x5 + x7 + x8 + x9 + x5sqrt + x7sqrt +
##     x8sqrt + x9sqrt, data = base)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89864 -0.79960 -0.00192  0.83311  3.02275
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5766526  0.6272645   0.919 0.358088
## x2          -0.2304361  0.0583390  -3.950 8.20e-05 ***
## x4          -0.0705862  0.0040770 -17.313  < 2e-16 ***
## x5          -0.3489673  0.0669042  -5.216 2.10e-07 ***
## x7           0.3784392  0.1313170   2.882 0.004013 **
## x8           0.2842105  0.1277121   2.225 0.026212 *
## x9           1.1409400  0.1276411   8.939  < 2e-16 ***
## x5sqrt       0.0090194  0.0014709   6.132 1.13e-09 ***
## x7sqrt      -0.0056810  0.0015116  -3.758 0.000178 ***
## x8sqrt      -0.0030764  0.0012017  -2.560 0.010571 *
## x9sqrt      -0.0108184  0.0006362 -17.004  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9876 on 1414 degrees of freedom
## Multiple R-squared:  0.9817, Adjusted R-squared:  0.9816
## F-statistic:  7579 on 10 and 1414 DF,  p-value: < 2.2e-16
```