Exercicio4.R

joaopaulodecker

2025-10-14

```r
# A - Amostra
base = read.csv2("selecao.csv")
set.seed(07052006)
base1 = base[sample(nrow(base), 500),]



# B - Seleção de variáveis - método backward
mc=lm( y ~ ., data=base1)
backward_model=step(mc, direction = "backward")
```

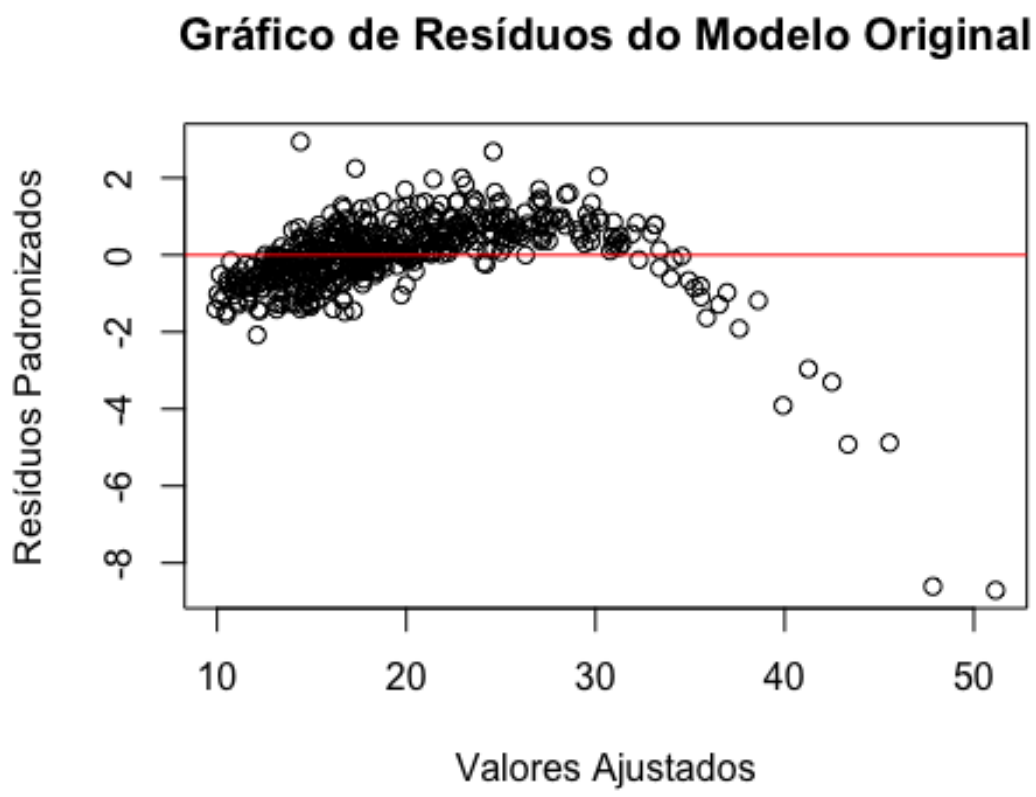
```
## Start:  AIC=956.74
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
##
##
## Step:  AIC=956.74
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x10
##
##        Df Sum of Sq    RSS     AIC
## <none>              3255.4  956.74
## - x2    1     13.45 3268.9  956.80
## - x3    1     13.49 3268.9  956.81
## - x5    1     32.10 3287.5  959.64
## - x1    1     45.68 3301.1  961.70
## - x6    1     54.75 3310.2  963.08
## - x4    1     66.66 3322.1  964.87
## - x8    1    467.93 3723.4 1021.89
## - x10   1    878.10 4133.5 1074.14
## - x7    1   2936.73 6192.2 1276.22
```

```r
summary(backward_model)
```

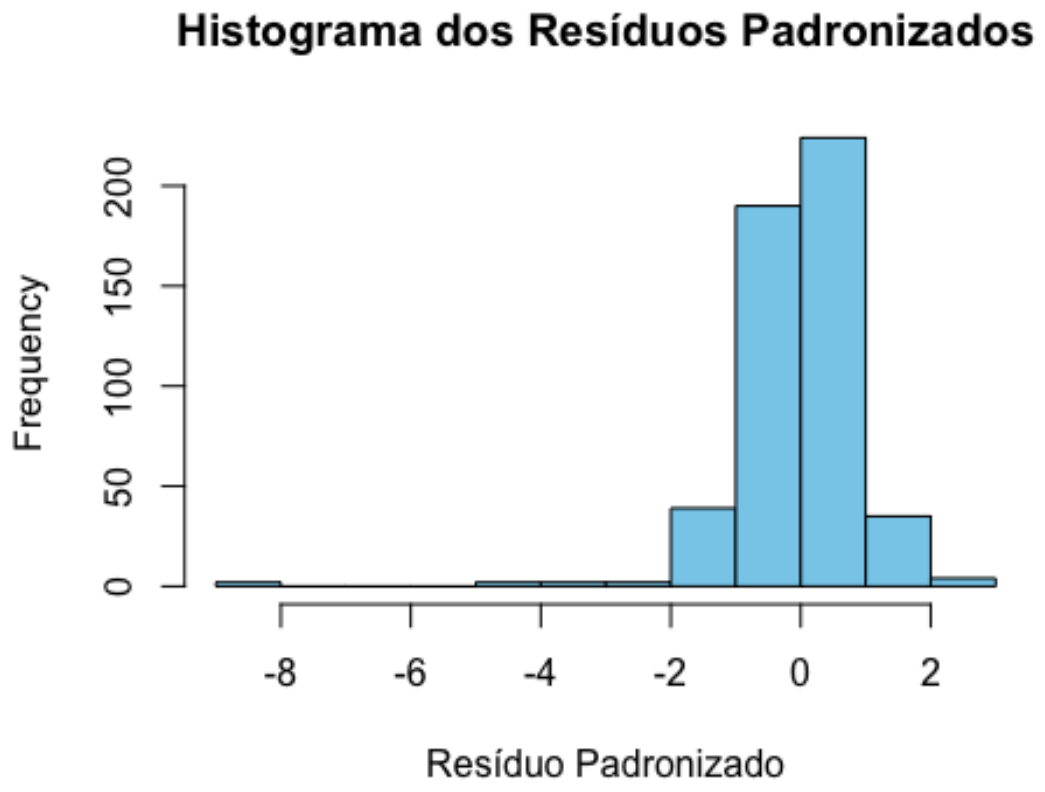
```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x10,
##     data = base1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.7933  -1.1019   0.1288   1.4379   7.2920
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.330105   1.755205  -0.188  0.85090
```

```
## x1          -0.001888   0.000720  -2.622  0.00901 **
## x2           0.397672   0.279450   1.423  0.15536
## x3          -0.001667   0.001169  -1.425  0.15473
## x4          -0.078398   0.024751  -3.167  0.00163 **
## x5           0.169575   0.077149   2.198  0.02841 *
## x6           0.086940   0.030286   2.871  0.00427 **
## x7           0.725019   0.034485  21.024  < 2e-16 ***
## x8           0.285508   0.034020   8.392 5.14e-16 ***
## x10          2.693361   0.234277  11.496  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.578 on 490 degrees of freedom
## Multiple R-squared:  0.8696, Adjusted R-squared:  0.8672
## F-statistic: 363.1 on 9 and 490 DF,  p-value: < 2.2e-16
```

```r
# C - Análise de resíduos
plot(fitted(backward_model), rstandard(backward_model),
     main = "Gráfico de Resíduos do Modelo Original",
     xlab = "Valores Ajustados",
     ylab = "Resíduos Padronizados")
abline(h = 0, col = "red")
```



```r
# Histograma dos resíduos
hist(rstandard(backward_model),
     main = "Histograma dos Resíduos Padronizados",
     xlab = "Resíduo Padronizado",
     col = "skyblue")
```




```r
# foi verificada a possibilidade de variáveis quadráticas
# D - backward com variáveis quadráticas
mcq = lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 +
           I(x1^2) + I(x2^2) + I(x3^2) + I(x4^2) + I(x5^2) +
           I(x6^2) + I(x7^2) + I(x8^2) + I(x9^2) + I(x10^2),
         data = base1)

backward_model_quad = step(mcq, direction = "backward")
```

```
## Start:  AIC=-806.71
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + I(x1^2) +
##     I(x2^2) + I(x3^2) + I(x4^2) + I(x5^2) + I(x6^2) + I(x7^2) +
##     I(x8^2) + I(x9^2) + I(x10^2)
##
##
## Step:  AIC=-806.71
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + I(x1^2) +
```

```
##     I(x2^2) + I(x3^2) + I(x4^2) + I(x5^2) + I(x6^2) + I(x7^2) +
##     I(x8^2) + I(x9^2)
##
##
## Step:  AIC=-806.71
## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x10 + I(x1^2) + I(x2^2) +
##     I(x3^2) + I(x4^2) + I(x5^2) + I(x6^2) + I(x7^2) + I(x8^2) +
##     I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - x6       1      0.00   92.31 -808.71
## - I(x1^2)  1      0.00   92.31 -808.71
## - I(x6^2)  1      0.00   92.31 -808.70
## - I(x4^2)  1      0.03   92.35 -808.53
## - I(x5^2)  1      0.04   92.36 -808.47
## - I(x8^2)  1      0.12   92.43 -808.08
## - x2       1      0.29   92.60 -807.15
## <none>                   92.31 -806.71
## - I(x2^2)  1      0.48   92.79 -806.13
## - x1       1      0.51   92.82 -805.98
## - I(x3^2)  1      0.61   92.92 -805.41
## - x3       1      0.63   92.94 -805.30
## - x5       1      0.83   93.15 -804.21
## - I(x7^2)  1      1.31   93.62 -801.66
## - x4       1      5.56   97.87 -779.46
## - I(x9^2)  1     67.97  160.29 -532.83
## - x10      1    361.78  454.09  -12.15
## - x7       1   1262.35 1354.66  534.35
## - x8       1   1417.94 1510.25  588.71
##
## Step:  AIC=-808.71
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x1^2) + I(x2^2) +
##     I(x3^2) + I(x4^2) + I(x5^2) + I(x6^2) + I(x7^2) + I(x8^2) +
##     I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - I(x1^2)  1      0.00   92.31 -810.71
## - I(x5^2)  1      0.05   92.36 -810.46
## - I(x4^2)  1      0.09   92.40 -810.23
## - I(x8^2)  1      0.12   92.43 -810.08
## - I(x6^2)  1      0.13   92.44 -810.03
## - x2       1      0.29   92.60 -809.15
## <none>                   92.31 -808.71
## - I(x2^2)  1      0.48   92.79 -808.13
## - x1       1      0.51   92.82 -807.98
## - I(x3^2)  1      0.61   92.93 -807.39
## - x3       1      0.63   92.95 -807.29
## - x5       1      0.84   93.16 -806.16
## - I(x7^2)  1      1.31   93.63 -803.66
## - x4       1     13.63  105.94 -741.88
```

```
## - I(x9^2)  1     68.32  160.63 -533.75
## - x10      1    362.03  454.34  -13.88
## - x7       1   1313.93 1406.24  551.03
## - x8       1   1420.96 1513.28  587.71
##
## Step:  AIC=-810.71
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) +
##     I(x4^2) + I(x5^2) + I(x6^2) + I(x7^2) + I(x8^2) + I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - I(x5^2)  1      0.05   92.36 -812.46
## - I(x4^2)  1      0.09   92.40 -812.23
## - I(x8^2)  1      0.12   92.43 -812.07
## - I(x6^2)  1      0.13   92.44 -812.02
## - x2       1      0.29   92.61 -811.13
## <none>                    92.31 -810.71
## - I(x2^2)  1      0.49   92.80 -810.08
## - I(x3^2)  1      0.61   92.93 -809.39
## - x3       1      0.63   92.95 -809.29
## - x5       1      0.85   93.16 -808.15
## - I(x7^2)  1      1.31   93.63 -805.65
## - x1       1      2.45   94.76 -799.61
## - x4       1     13.71  106.02 -743.47
## - I(x9^2)  1     68.32  160.64 -535.73
## - x10      1    362.08  454.40  -15.82
## - x7       1   1314.53 1406.84  549.25
## - x8       1   1421.40 1513.71  585.86
##
## Step:  AIC=-812.46
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) +
##     I(x4^2) + I(x6^2) + I(x7^2) + I(x8^2) + I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - I(x4^2)  1      0.05   92.41 -814.21
## - I(x6^2)  1      0.09   92.45 -813.98
## - I(x8^2)  1      0.12   92.48 -813.83
## - x2       1      0.31   92.67 -812.78
## <none>                    92.36 -812.46
## - I(x2^2)  1      0.51   92.87 -811.73
## - I(x3^2)  1      0.64   93.00 -811.00
## - x3       1      0.65   93.01 -810.93
## - I(x7^2)  1      1.30   93.66 -807.48
## - x1       1      2.44   94.80 -801.42
## - x5       1      9.89  102.25 -763.59
## - x4       1     16.30  108.66 -733.21
## - I(x9^2)  1     69.10  161.46 -535.18
## - x10      1    366.83  459.20  -12.57
## - x7       1   1322.39 1414.75  550.05
## - x8       1   1476.36 1568.72  601.70
##
```

```
## Step:  AIC=-814.21
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) +
##     I(x6^2) + I(x7^2) + I(x8^2) + I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - I(x6^2)  1      0.09   92.49 -815.73
## - I(x8^2)  1      0.10   92.51 -815.67
## - x2       1      0.33   92.73 -814.43
## <none>                   92.41 -814.21
## - I(x2^2)  1      0.52   92.93 -813.38
## - I(x3^2)  1      0.63   93.03 -812.82
## - x3       1      0.64   93.05 -812.75
## - I(x7^2)  1      1.29   93.70 -809.28
## - x1       1      2.63   95.04 -802.17
## - x5       1     10.04  102.45 -764.62
## - x4       1     58.99  151.40 -569.34
## - I(x9^2)  1     69.35  161.76 -536.25
## - x10      1    371.34  463.75   -9.64
## - x7       1   1322.36 1414.76  548.05
## - x8       1   1534.69 1627.10  617.97
##
## Step:  AIC=-815.73
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) +
##     I(x7^2) + I(x8^2) + I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - I(x8^2)  1      0.07   92.57 -817.34
## - x2       1      0.33   92.82 -815.97
## <none>                   92.49 -815.73
## - I(x2^2)  1      0.52   93.01 -814.94
## - I(x3^2)  1      0.60   93.10 -814.48
## - x3       1      0.62   93.11 -814.39
## - I(x7^2)  1      1.23   93.72 -811.13
## - x1       1      2.64   95.13 -803.68
## - x5       1     11.14  103.64 -760.85
## - I(x9^2)  1     71.38  163.88 -531.75
## - x4       1    133.38  225.87 -371.31
## - x10      1    371.43  463.92  -11.45
## - x7       1   1333.10 1425.59  549.87
## - x8       1   1618.39 1710.88  641.08
##
## Step:  AIC=-817.34
## y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) +
##     I(x7^2) + I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## - x2       1      0.33   92.90 -817.53
## <none>                   92.57 -817.34
## - I(x2^2)  1      0.52   93.09 -816.54
## - I(x3^2)  1      0.62   93.19 -816.01
```

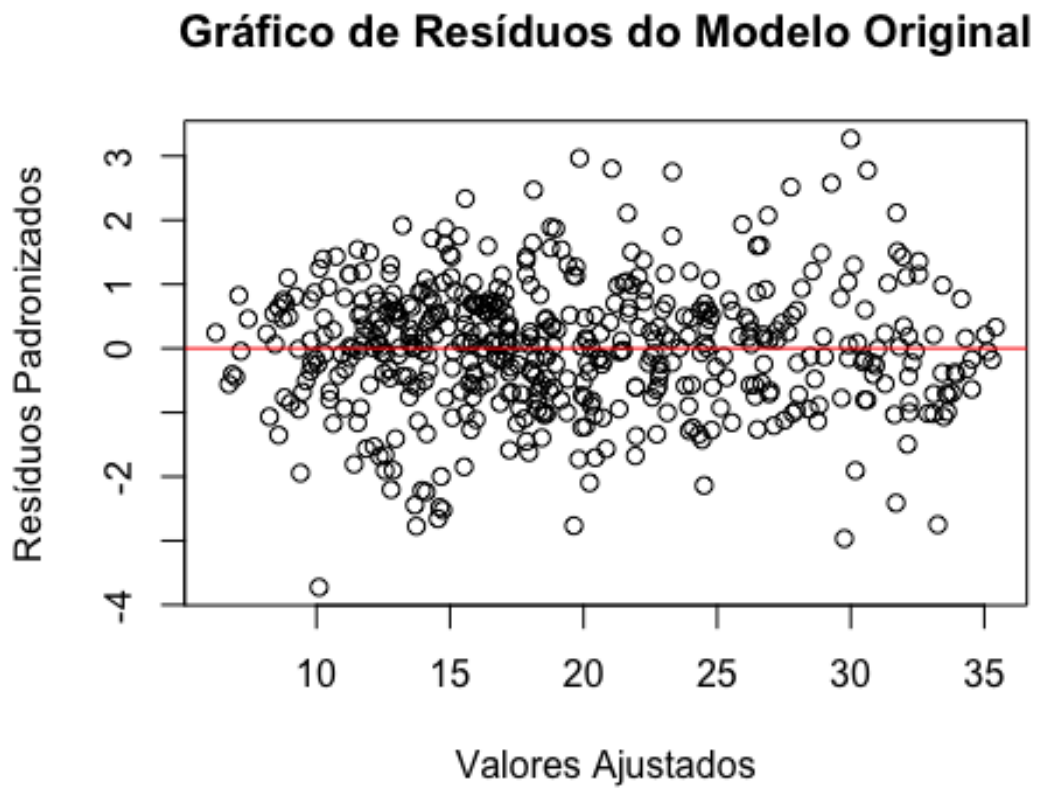
```
## - x3       1      0.63   93.20 -815.93
## - x1       1      2.73   95.30 -804.79
## - I(x7^2)  1      2.76   95.32 -804.67
## - x5       1     11.48  104.05 -760.87
## - x4       1    133.33  225.90 -373.27
## - x10      1    372.04  464.61  -12.71
## - I(x9^2)  1    831.47  924.04  331.07
## - x7       1   1377.09 1469.66  563.09
## - x8       1   1618.39 1710.96  639.10
##
## Step:  AIC=-817.53
## y ~ x1 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) + I(x3^2) + I(x7^2) +
##     I(x9^2)
##
##           Df Sum of Sq     RSS     AIC
## <none>                   92.90 -817.53
## - I(x2^2)  1      0.39   93.29 -817.44
## - I(x3^2)  1      0.60   93.50 -816.34
## - x3       1      0.62   93.52 -816.20
## - x1       1      2.40   95.31 -806.76
## - I(x7^2)  1      2.77   95.67 -804.87
## - x5       1     11.52  104.42 -761.10
## - x4       1    133.66  226.56 -373.80
## - x10      1    372.44  465.34  -13.92
## - I(x9^2)  1    832.98  925.88  330.07
## - x7       1   1377.28 1470.18  561.27
## - x8       1   1622.72 1715.63  638.46
```

```
summary(backward_model_quad)
```

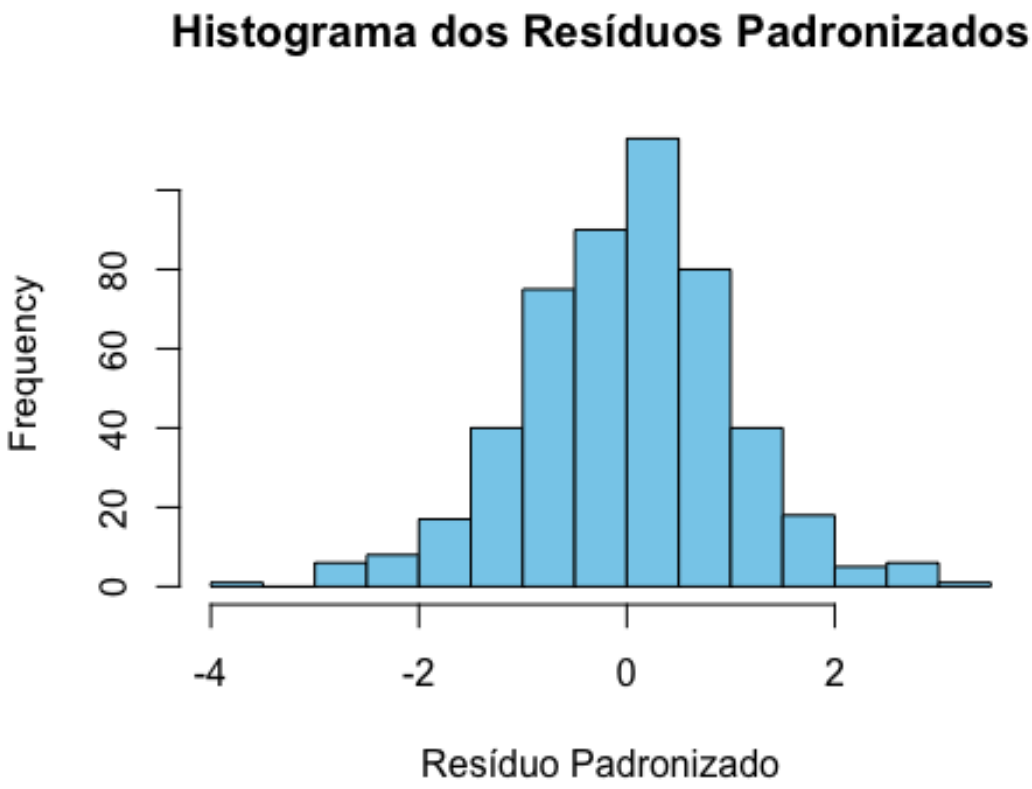
```
##
## Call:
## lm(formula = y ~ x1 + x3 + x4 + x5 + x7 + x8 + x10 + I(x2^2) +
##     I(x3^2) + I(x7^2) + I(x9^2), data = base1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61814 -0.26235  0.01461  0.25871  1.39703
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.160e+00  1.855e-01 -22.428  < 2e-16 ***
## x1          -4.248e-04  1.195e-04  -3.554 0.000416 ***
## x3           1.286e-03  7.123e-04   1.806 0.071604 .
## x4          -7.707e-02  2.909e-03 -26.497  < 2e-16 ***
## x5           9.882e-02  1.270e-02   7.778 4.41e-14 ***
## x7           1.435e+00  1.687e-02  85.057  < 2e-16 ***
## x8           1.336e+00  1.447e-02  92.325  < 2e-16 ***
## x10          1.781e+00  4.026e-02  44.231  < 2e-16 ***
## I(x2^2)     -1.571e-02  1.098e-02  -1.430 0.153298
```

```
## I(x3^2)     -2.745e-06  1.552e-06  -1.768 0.077638 .
## I(x7^2)     -2.657e-03  6.971e-04  -3.811 0.000156 ***
## I(x9^2)     -1.146e-02  1.733e-04 -66.148  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4363 on 488 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9962
## F-statistic: 1.188e+04 on 11 and 488 DF,  p-value: < 2.2e-16
```

```r
#  Análise de resíduos
plot(fitted(backward_model_quad), rstandard(backward_model_quad),
     main = "Gráfico de Resíduos do Modelo Original",
     xlab = "Valores Ajustados",
     ylab = "Resíduos Padronizados")
abline(h = 0, col = "red")
```




```r
# Histograma dos resíduos
hist(rstandard(backward_model_quad),
     main = "Histograma dos Resíduos Padronizados",
     xlab = "Resíduo Padronizado",
     col = "skyblue")
```


Histograma dos Resíduos Padronizados
Frequency
0
20
40
60
80
-4
-2
0
2
Resíduo Padronizado