

# Técnicas Estatísticas de Predição

## Exercício 3

João Pedro Gentil da Silveira

---

### Questões

A) Retire uma amostra de 1.000 observações utilizando a sintaxe com seed.

R:

```
# Leitura da base de dados
base = read.csv2("selecao.csv", dec=".")

# Tira a amostra de tamanho 1000 da base de dados
set.seed(29072003)
amostra = base[sample(nrow(base), 1000),]
```

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

R:

```
# Cria um modelo zero
modelo_zero_1 = lm(y ~ 1, data=amostra)

# Constrói um modelo adequado
modelo_adequado_1 = step(modelo_zero_1, list(lower = ~ 1,
                                             upper = ~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10),
                          direction="forward")

## Start:  AIC=3947.66
## y ~ 1
##
##           Df Sum of Sq  RSS   AIC
## + x7       1     35999 15711 2758.4
## + x9       1     35718 15992 2776.1
## + x8       1     28728 22982 3138.7
## + x5       1     22941 28769 3363.3
## + x6       1     22811 28899 3367.8
## + x4       1     12388 39323 3675.8
## + x10      1      5174 46536 3844.2
## + x2       1       328 51382 3943.3
## <none>          51710 3947.7
## + x3       1        83 51628 3948.1
## + x1       1        42 51668 3948.8
##
## Step:  AIC=2758.37
## y ~ x7
##
```

```

##          Df Sum of Sq  RSS    AIC
## + x10    1   1929.60 13782 2629.3
## + x9      1    948.99 14762 2698.1
## + x8      1    924.17 14787 2699.8
## + x6      1    730.05 14981 2712.8
## + x5      1    423.28 15288 2733.1
## + x4      1     85.79 15625 2754.9
## + x1      1     42.26 15669 2757.7
## <none>                15711 2758.4
## + x2      1      3.51 15708 2760.2
## + x3      1      0.52 15711 2760.3
##
## Step: AIC=2629.33
## y ~ x7 + x10
##
##          Df Sum of Sq  RSS    AIC
## + x9      1    886.13 12895 2564.9
## + x8      1    866.27 12915 2566.4
## + x6      1    790.08 12992 2572.3
## + x5      1    546.05 13236 2590.9
## + x4      1    136.66 13645 2621.4
## <none>                13782 2629.3
## + x1      1     18.49 13763 2630.0
## + x3      1      1.64 13780 2631.2
## + x2      1      0.00 13782 2631.3
##
## Step: AIC=2564.87
## y ~ x7 + x10 + x9
##
##          Df Sum of Sq  RSS    AIC
## + x6      1    224.381 12671 2549.3
## + x5      1     83.500 12812 2560.4
## <none>                12895 2564.9
## + x8      1     25.121 12870 2564.9
## + x3      1     11.220 12884 2566.0
## + x1      1      9.690 12886 2566.1
## + x4      1      3.913 12892 2566.6
## + x2      1      2.532 12893 2566.7
##
## Step: AIC=2549.32
## y ~ x7 + x10 + x9 + x6
##
##          Df Sum of Sq  RSS    AIC
## + x4      1     589.77 12081 2503.7
## + x1      1      29.20 12642 2549.0
## + x8      1      26.14 12645 2549.2
## <none>                12671 2549.3
## + x3      1      24.33 12647 2549.4
## + x2      1       1.25 12670 2551.2
## + x5      1       0.78 12670 2551.3
##
## Step: AIC=2503.66
## y ~ x7 + x10 + x9 + x6 + x4

```

```
##
##      Df Sum of Sq  RSS   AIC
## + x5    1    362.09 11719 2475.2
## + x8    1     25.64 12056 2503.5
## <none>                12081 2503.7
## + x1    1     17.55 12064 2504.2
## + x3    1     12.79 12068 2504.6
## + x2    1      1.90 12079 2505.5
##
## Step: AIC=2475.23
## y ~ x7 + x10 + x9 + x6 + x4 + x5
##
##      Df Sum of Sq  RSS   AIC
## + x8    1    30.0528 11689 2474.7
## <none>                11719 2475.2
## + x1    1    10.7140 11708 2476.3
## + x3    1     6.3050 11713 2476.7
## + x2    1     1.1971 11718 2477.1
##
## Step: AIC=2474.66
## y ~ x7 + x10 + x9 + x6 + x4 + x5 + x8
##
##      Df Sum of Sq  RSS   AIC
## <none>                11689 2474.7
## + x1    1    11.3296 11678 2475.7
## + x3    1     6.8461 11682 2476.1
## + x2    1     1.2741 11688 2476.6

summary(modelo_adequado_1)

##
## Call:
## lm(formula = y ~ x7 + x10 + x9 + x6 + x4 + x5 + x8, data = amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.198  -1.625   0.027   1.777  14.819
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.98571    1.59273  -5.642 2.20e-08 ***
## x7           -0.21528    0.44315  -0.486  0.6272
## x10          2.75326    0.22090  12.464 < 2e-16 ***
## x9           0.80782    0.44107   1.832  0.0673 .
## x6           0.26605    0.03006   8.851 < 2e-16 ***
## x4          -0.21388    0.02376  -9.002 < 2e-16 ***
## x5           0.43947    0.07880   5.577 3.15e-08 ***
## x8          -0.70437    0.44106  -1.597  0.1106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.433 on 992 degrees of freedom
```

```
## Multiple R-squared:  0.7739, Adjusted R-squared:  0.7724  
## F-statistic: 485.2 on 7 and 992 DF,  p-value: < 2.2e-16
```

**C)** 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 (se necessário fazer alguma transformação ou acrescentar algum termo polinomial no modelo).

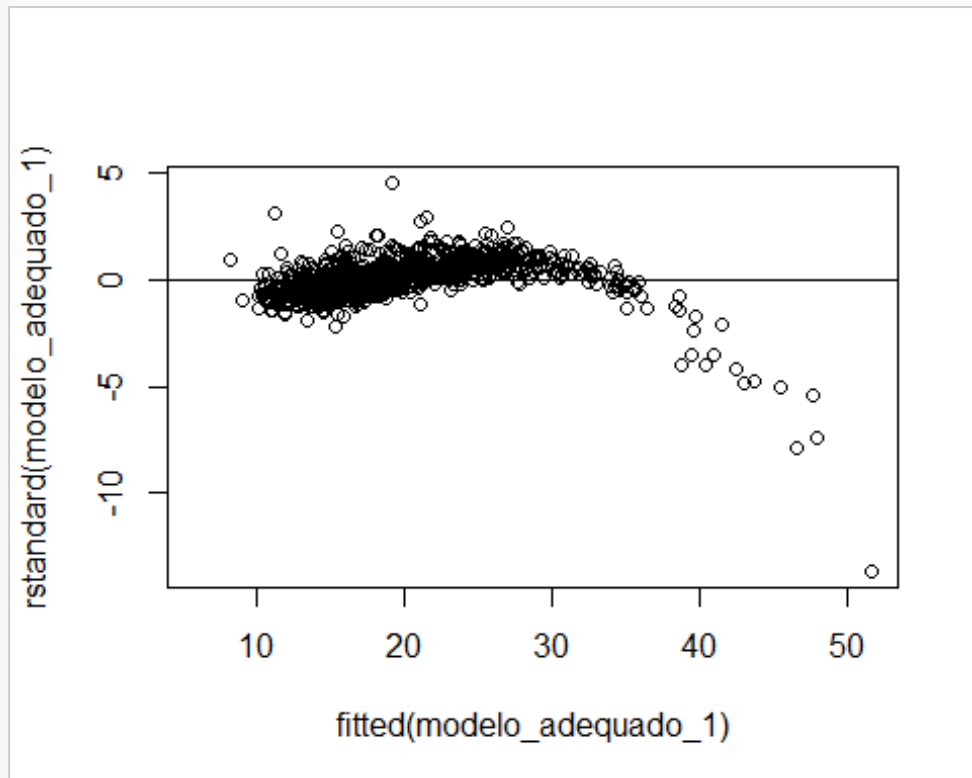
**R:** Feita a análise de resíduos do modelo obtido até então, nota-se que o seu gráfico de resíduos apresenta uma variância tendenciosa e não aleatória, livre de padrões; além disso, seu valor de  $R^2$  não é muito elevado. Portanto, cresce-se a necessidade de fazer transformações no modelo para que essas defasagens sejam supridas.

A primeira tentativa de ajuste foi optar por uma função logarítmica que, justamente, usufruía do valor do log do y predito. Seus resultados não foram efetivos o bastante, resultando em um gráfico com variância tendenciosa e valor de  $R^2$  inferior quando comparado ao anterior.

A segunda tentativa de ajuste foi optar por uma função logarítmica que usufruiu não só do valor do log do y predito, como também da variável independente x9. Nesse caso, o modelo teve ajuste muito significativo, com seu valor de  $R^2$  se aproximando de 0,98 e com um gráfico de resíduos muito mais disperso e despadronizado.

Vale dizer que outras tentativas de obter um modelo ainda melhor foram feitas, mas foram omitidas da sintaxe para torná-lo mais sucinto, uma vez que todas elas não se mostraram melhores do que a última tentativa apresentada aqui.

```
# Análise de resíduos do modelo  
plot(fitted(modelo_adequado_1), rstandard(modelo_adequado_1))  
abline(0,0)
```



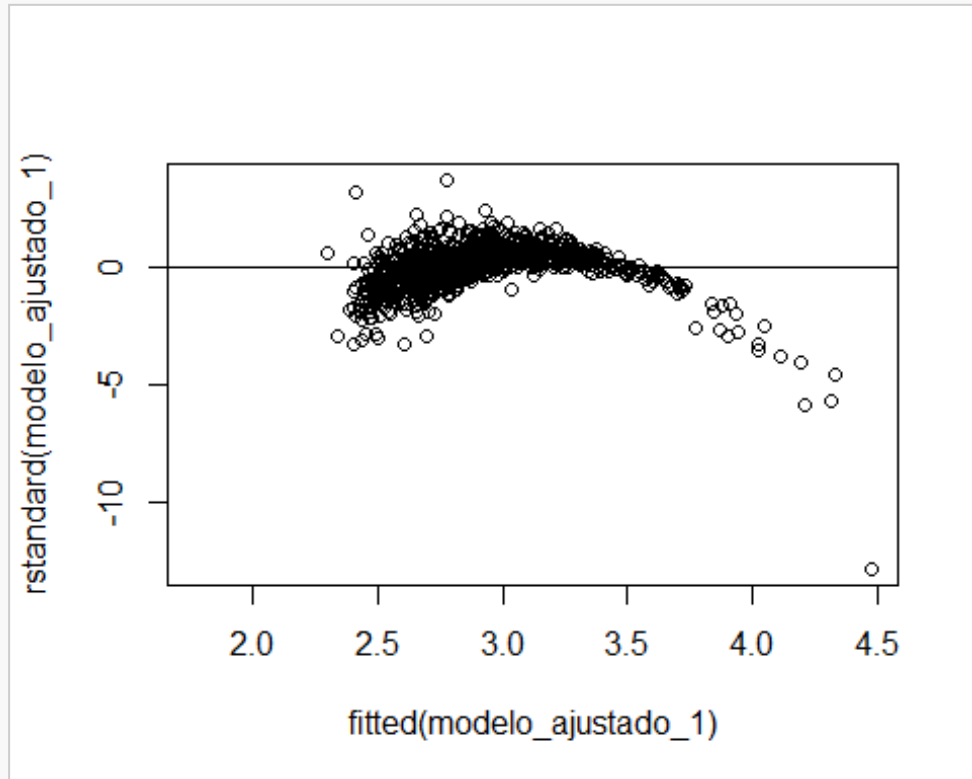
#### # Ajustes de modelos

```
modelo_ajustado_1 = lm(log(y) ~ x7+x10+x9+x6+x4+x5+x8, data=amostra)
summary(modelo_ajustado_1)
```

```
##
## Call:
## lm(formula = log(y) ~ x7 + x10 + x9 + x6 + x4 + x5 + x8, data =
## amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61739 -0.08234  0.03084  0.12060  0.75307
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.422722   0.097585  14.579  < 2e-16 ***
## x7           -0.060074   0.027152  -2.213  0.027156 *
## x10           0.175972   0.013534  13.002  < 2e-16 ***
## x9            0.093325   0.027024   3.453  0.000577 ***
## x6            0.013482   0.001842   7.320  5.12e-13 ***
## x4           -0.011465   0.001456  -7.876  8.87e-15 ***
## x5            0.024379   0.004828   5.049  5.27e-07 ***
## x8           -0.092176   0.027023  -3.411  0.000673 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2103 on 992 degrees of freedom
```

```
## Multiple R-squared:  0.7065, Adjusted R-squared:  0.7044
## F-statistic: 341.1 on 7 and 992 DF,  p-value: < 2.2e-16

plot(fitted(modelo_ajustado_1), rstandard(modelo_ajustado_1))
abline(0,0)
```

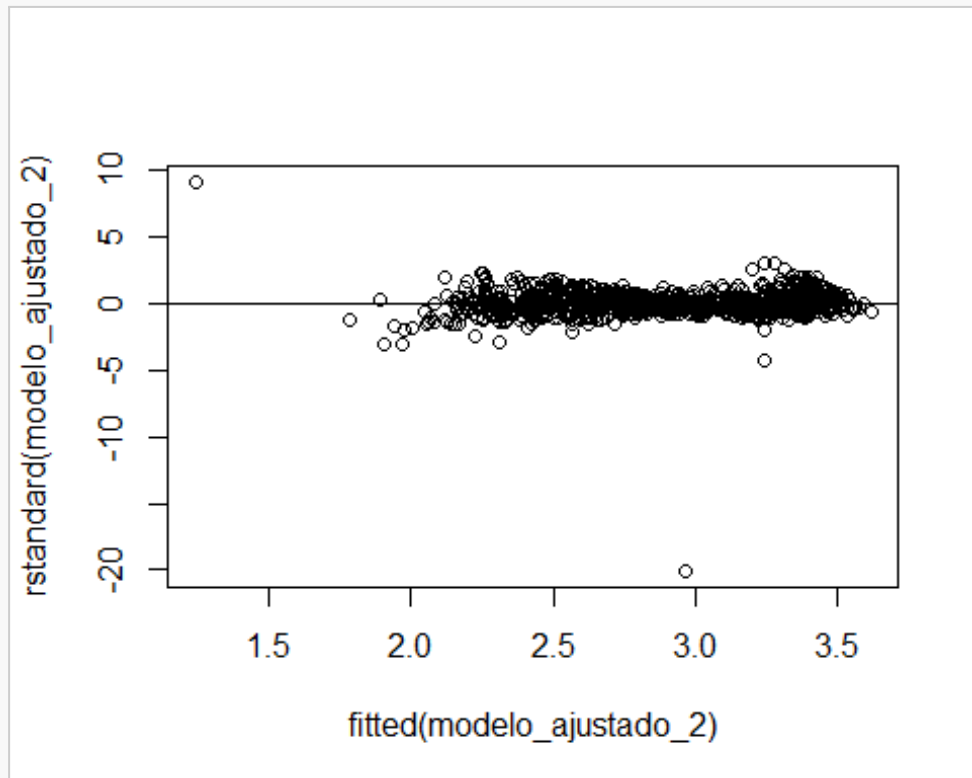


```
modelo_ajustado_2 = lm(log(y) ~ x7+x10+log(x9)+x6+x4+x5+x8,
data=amostra)
summary(modelo_ajustado_2)

##
## Call:
## lm(formula = log(y) ~ x7 + x10 + log(x9) + x6 + x4 + x5 + x8,
##     data = amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11041 -0.02495 -0.00321  0.02338  0.52821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.3521830   0.0374825  -36.075  < 2e-16 ***
## x7           -0.0311243   0.0008071  -38.562  < 2e-16 ***
## x10           0.1078133   0.0038565   27.956  < 2e-16 ***
## log(x9)       1.6232358   0.0149975  108.234  < 2e-16 ***
## x6            0.0002289   0.0005321    0.430    0.667
## x4           -0.0054487   0.0004129  -13.195  < 2e-16 ***
## x5            0.0090698   0.0013645    6.647 4.92e-11 ***
## x8           -0.0295803   0.0006038  -48.994  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05912 on 992 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.9766
## F-statistic: 5970 on 7 and 992 DF,  p-value: < 2.2e-16

plot(fitted(modelo_ajustado_2), rstandard(modelo_ajustado_2))
abline(0,0)
```



**D)** Utilize novamente o método de seleção de variáveis “forward” e refaça a análise começando do item (b).

**R:** A partir da interpretação de que a questão exigia que fosse feito o método de seleção de variáveis para as condições do novo modelo obtido na questão anterior, construiu-se um novo modelo adequado, usando do valor do logaritmo do y predito e da variável independente x9. Nesse processo, encontrou-se um mínimo benefício nos valores de  $R^2$  e na dispersão mostrada pelo gráfico de resíduos ao se retirar a variável x6 da predição.

```
# Constrói um novo modelo adequado
modelo_zero_2 = lm(log(y) ~ 1, data=amostra)
modelo_adequado_2 = step(modelo_zero_2, list(lower = ~ 1,
                                              upper = ~
x7+x10+log(x9)+x6+x4+x5+x8),
                          direction="forward")
```

```

## Start:  AIC=-1898.38
## log(y) ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + log(x9)  1   125.213   24.299 -3713.3
## + x7       1    93.211   56.301 -2873.0
## + x8       1    68.513   80.999 -2509.3
## + x5       1    55.810   93.702 -2363.6
## + x6       1    54.978   94.534 -2354.8
## + x4       1    28.678  120.834 -2109.3
## + x10      1    17.764  131.748 -2022.9
## <none>                149.512 -1898.4
##
## Step:  AIC=-3713.33
## log(y) ~ log(x9)
##
##           Df Sum of Sq    RSS    AIC
## + x8       1    10.9783  13.320 -4312.5
## + x10      1     4.6077  19.691 -3921.6
## + x4       1     4.4198  19.879 -3912.1
## + x5       1     4.1093  20.189 -3896.6
## + x6       1     3.8142  20.485 -3882.1
## + x7       1     3.6346  20.664 -3873.4
## <none>                24.299 -3713.3
##
## Step:  AIC=-4312.45
## log(y) ~ log(x9) + x8
##
##           Df Sum of Sq    RSS    AIC
## + x7       1     5.1967   8.1238 -4805.0
## + x10      1     3.6793   9.6412 -4633.7
## + x4       1     1.3306  11.9899 -4415.7
## + x6       1     0.9316  12.3889 -4383.0
## + x5       1     0.8064  12.5141 -4372.9
## <none>                13.3205 -4312.5
##
## Step:  AIC=-4804.96
## log(y) ~ log(x9) + x8 + x7
##
##           Df Sum of Sq    RSS    AIC
## + x10      1     3.07737  5.0464 -5279.1
## + x4       1     1.84001  6.2838 -5059.8
## + x6       1     1.25807  6.8657 -4971.2
## + x5       1     0.64829  7.4755 -4886.1
## <none>                8.1238 -4805.0
##
## Step:  AIC=-5279.07
## log(y) ~ log(x9) + x8 + x7 + x10
##
##           Df Sum of Sq    RSS    AIC
## + x4       1     1.42444  3.6220 -5608.7
## + x6       1     0.96233  4.0841 -5488.7
## + x5       1     0.37509  4.6713 -5354.3

```

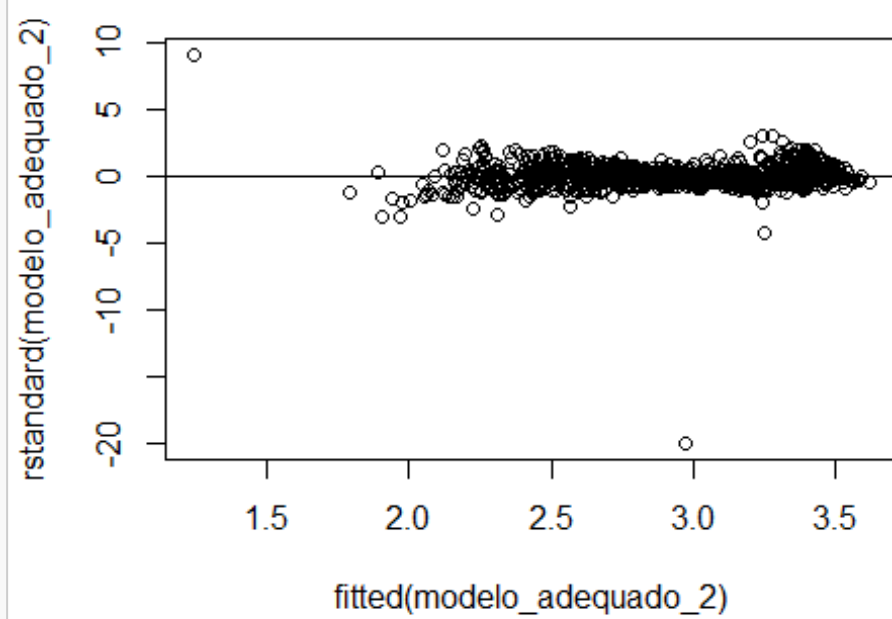


```
## <none>          5.0464 -5279.1
##
## Step: AIC=-5608.73
## log(y) ~ log(x9) + x8 + x7 + x10 + x4
##
##           Df Sum of Sq    RSS    AIC
## + x5       1  0.154421  3.4676 -5650.3
## <none>          3.6220 -5608.7
## + x6       1  0.000645  3.6214 -5606.9
##
## Step: AIC=-5650.3
## log(y) ~ log(x9) + x8 + x7 + x10 + x4 + x5
##
##           Df Sum of Sq    RSS    AIC
## <none>          3.4676 -5650.3
## + x6       1 0.00064646  3.4669 -5648.5

summary(modelo_adequado_2)

##
## Call:
## lm(formula = log(y) ~ log(x9) + x8 + x7 + x10 + x4 + x5, data =
amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11221 -0.02503 -0.00346  0.02302  0.52897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.3449384  0.0334706  -40.18  < 2e-16 ***
## log(x9)      1.6247313  0.0145829  111.41  < 2e-16 ***
## x8          -0.0295436  0.0005975  -49.45  < 2e-16 ***
## x7          -0.0311160  0.0008066  -38.58  < 2e-16 ***
## x10          0.1077881  0.0038545   27.96  < 2e-16 ***
## x4          -0.0053209  0.0002866  -18.57  < 2e-16 ***
## x5           0.0089942  0.0013525    6.65 4.83e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05909 on 993 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.9767
## F-statistic: 6970 on 6 and 993 DF, p-value: < 2.2e-16

plot(fitted(modelo_adequado_2), rstandard(modelo_adequado_2))
abline(0,0)
```



## Sintaxe Completa

```
# Leitura da base de dados
base = read.csv2("selecao.csv", dec=".")

# Tira a amostra de tamanho 1000 da base de dados
set.seed(29072003)
amostra = base[sample(nrow(base), 1000),]

# Cria um modelo zero
modelo_zero_1 = lm(y ~ 1, data=amostra)

# Constrói um modelo adequado
modelo_adequado_1 = step(modelo_zero_1, list(lower = ~ 1,
      upper = ~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10),
      direction="forward")

## Start:  AIC=3947.66
## y ~ 1
##
##           Df Sum of Sq  RSS    AIC
## + x7       1     35999 15711 2758.4
## + x9       1     35718 15992 2776.1
## + x8       1     28728 22982 3138.7
## + x5       1     22941 28769 3363.3
## + x6       1     22811 28899 3367.8
## + x4       1     12388 39323 3675.8
## + x10      1       5174 46536 3844.2
## + x2       1        328 51382 3943.3
## <none>                 51710 3947.7
## + x3       1         83 51628 3948.1
## + x1       1         42 51668 3948.8
##
## Step:  AIC=2758.37
## y ~ x7
##
##           Df Sum of Sq  RSS    AIC
## + x10      1    1929.60 13782 2629.3
## + x9       1     948.99 14762 2698.1
## + x8       1     924.17 14787 2699.8
## + x6       1     730.05 14981 2712.8
## + x5       1     423.28 15288 2733.1
## + x4       1      85.79 15625 2754.9
## + x1       1      42.26 15669 2757.7
## <none>                 15711 2758.4
## + x2       1        3.51 15708 2760.2
## + x3       1         0.52 15711 2760.3
##
## Step:  AIC=2629.33
## y ~ x7 + x10
##
##           Df Sum of Sq  RSS    AIC
## + x9       1     886.13 12895 2564.9
```

```

## + x8      1      866.27 12915 2566.4
## + x6      1      790.08 12992 2572.3
## + x5      1      546.05 13236 2590.9
## + x4      1      136.66 13645 2621.4
## <none>                13782 2629.3
## + x1      1       18.49 13763 2630.0
## + x3      1        1.64 13780 2631.2
## + x2      1         0.00 13782 2631.3
##
## Step: AIC=2564.87
## y ~ x7 + x10 + x9
##
##           Df Sum of Sq  RSS    AIC
## + x6      1   224.381 12671 2549.3
## + x5      1    83.500 12812 2560.4
## <none>                12895 2564.9
## + x8      1    25.121 12870 2564.9
## + x3      1    11.220 12884 2566.0
## + x1      1     9.690 12886 2566.1
## + x4      1     3.913 12892 2566.6
## + x2      1     2.532 12893 2566.7
##
## Step: AIC=2549.32
## y ~ x7 + x10 + x9 + x6
##
##           Df Sum of Sq  RSS    AIC
## + x4      1    589.77 12081 2503.7
## + x1      1     29.20 12642 2549.0
## + x8      1     26.14 12645 2549.2
## <none>                12671 2549.3
## + x3      1     24.33 12647 2549.4
## + x2      1      1.25 12670 2551.2
## + x5      1      0.78 12670 2551.3
##
## Step: AIC=2503.66
## y ~ x7 + x10 + x9 + x6 + x4
##
##           Df Sum of Sq  RSS    AIC
## + x5      1    362.09 11719 2475.2
## + x8      1     25.64 12056 2503.5
## <none>                12081 2503.7
## + x1      1     17.55 12064 2504.2
## + x3      1     12.79 12068 2504.6
## + x2      1      1.90 12079 2505.5
##
## Step: AIC=2475.23
## y ~ x7 + x10 + x9 + x6 + x4 + x5
##
##           Df Sum of Sq  RSS    AIC
## + x8      1    30.0528 11689 2474.7
## <none>                11719 2475.2
## + x1      1    10.7140 11708 2476.3
## + x3      1     6.3050 11713 2476.7

```

```

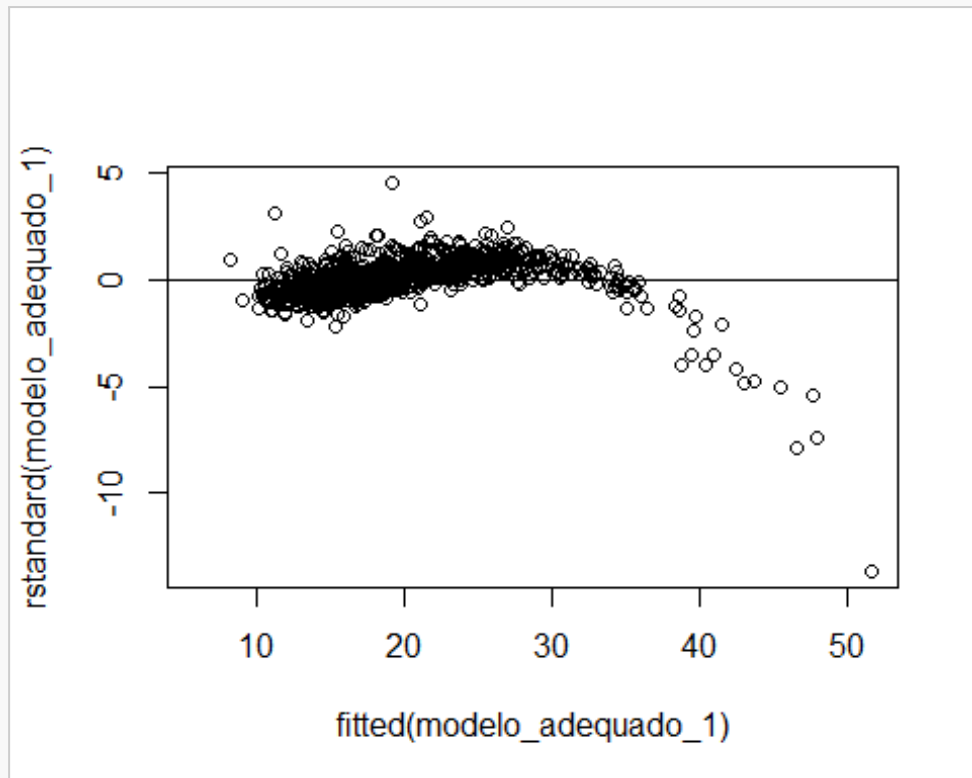
## + x2      1      1.1971 11718 2477.1
##
## Step: AIC=2474.66
## y ~ x7 + x10 + x9 + x6 + x4 + x5 + x8
##
##           Df Sum of Sq  RSS   AIC
## <none>                11689 2474.7
## + x1      1     11.3296 11678 2475.7
## + x3      1      6.8461 11682 2476.1
## + x2      1      1.2741 11688 2476.6

summary(modelo_adequado_1)

##
## Call:
## lm(formula = y ~ x7 + x10 + x9 + x6 + x4 + x5 + x8, data = amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.198  -1.625   0.027   1.777  14.819
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.98571     1.59273  -5.642 2.20e-08 ***
## x7           -0.21528     0.44315  -0.486  0.6272
## x10          2.75326     0.22090  12.464 < 2e-16 ***
## x9           0.80782     0.44107   1.832  0.0673 .
## x6           0.26605     0.03006   8.851 < 2e-16 ***
## x4          -0.21388     0.02376  -9.002 < 2e-16 ***
## x5           0.43947     0.07880   5.577 3.15e-08 ***
## x8          -0.70437     0.44106  -1.597  0.1106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.433 on 992 degrees of freedom
## Multiple R-squared:  0.7739, Adjusted R-squared:  0.7724
## F-statistic: 485.2 on 7 and 992 DF, p-value: < 2.2e-16

# Análise de resíduos do modelo
plot(fitted(modelo_adequado_1), rstandard(modelo_adequado_1))
abline(0,0)

```



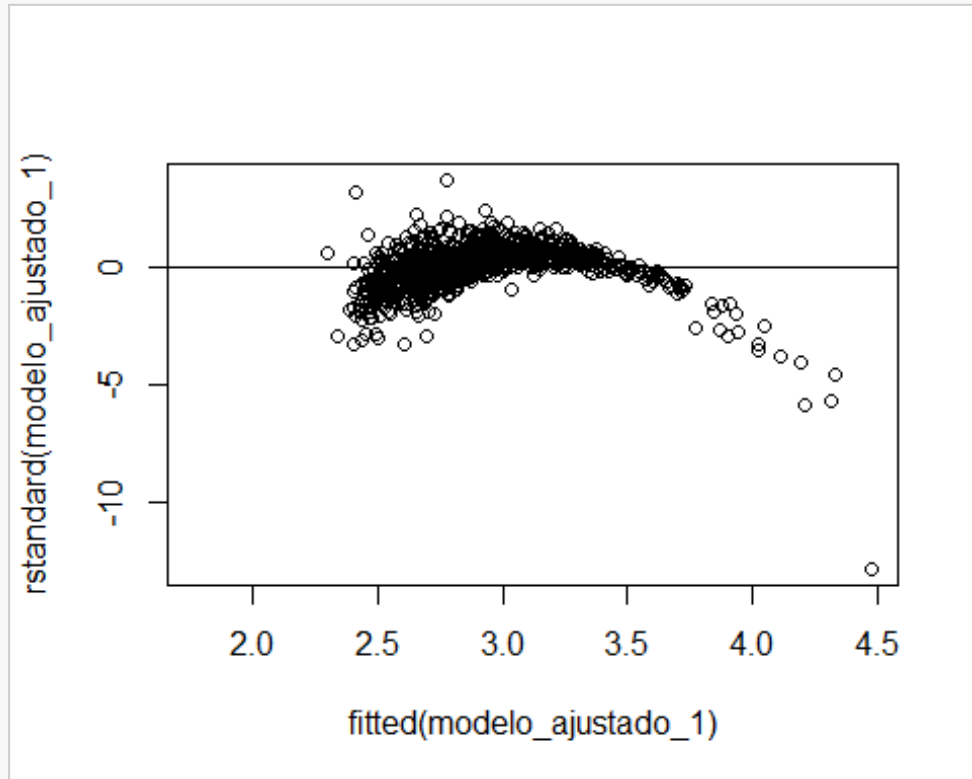
#### # Ajustes de modelos

```
modelo_ajustado_1 = lm(log(y) ~ x7+x10+x9+x6+x4+x5+x8, data=amostra)
summary(modelo_ajustado_1)
```

```
##
## Call:
## lm(formula = log(y) ~ x7 + x10 + x9 + x6 + x4 + x5 + x8, data =
## amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61739 -0.08234  0.03084  0.12060  0.75307
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.422722   0.097585  14.579  < 2e-16 ***
## x7           -0.060074   0.027152  -2.213  0.027156 *
## x10           0.175972   0.013534  13.002  < 2e-16 ***
## x9            0.093325   0.027024   3.453  0.000577 ***
## x6            0.013482   0.001842   7.320  5.12e-13 ***
## x4           -0.011465   0.001456  -7.876  8.87e-15 ***
## x5            0.024379   0.004828   5.049  5.27e-07 ***
## x8           -0.092176   0.027023  -3.411  0.000673 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2103 on 992 degrees of freedom
```

```
## Multiple R-squared:  0.7065, Adjusted R-squared:  0.7044
## F-statistic: 341.1 on 7 and 992 DF,  p-value: < 2.2e-16

plot(fitted(modelo_ajustado_1), rstandard(modelo_ajustado_1))
abline(0,0)
```

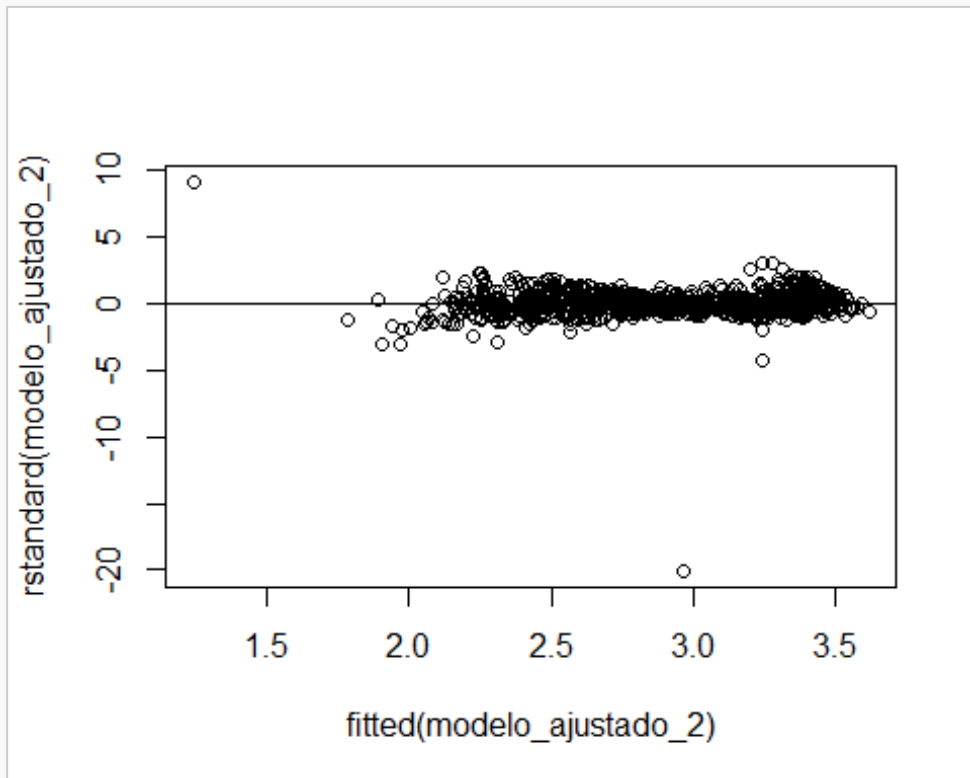


```
modelo_ajustado_2 = lm(log(y) ~ x7+x10+log(x9)+x6+x4+x5+x8,
data=amostra)
summary(modelo_ajustado_2)

##
## Call:
## lm(formula = log(y) ~ x7 + x10 + log(x9) + x6 + x4 + x5 + x8,
##     data = amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11041 -0.02495 -0.00321  0.02338  0.52821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.3521830   0.0374825  -36.075  < 2e-16 ***
## x7           -0.0311243   0.0008071  -38.562  < 2e-16 ***
## x10           0.1078133   0.0038565   27.956  < 2e-16 ***
## log(x9)       1.6232358   0.0149975  108.234  < 2e-16 ***
## x6            0.0002289   0.0005321    0.430    0.667
## x4           -0.0054487   0.0004129  -13.195  < 2e-16 ***
## x5            0.0090698   0.0013645    6.647 4.92e-11 ***
## x8           -0.0295803   0.0006038  -48.994  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05912 on 992 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.9766
## F-statistic: 5970 on 7 and 992 DF,  p-value: < 2.2e-16

plot(fitted(modelo_ajustado_2), rstandard(modelo_ajustado_2))
abline(0,0)
```



*# Constrói um novo modelo adequado*

```
modelo_zero_2 = lm(log(y) ~ 1, data=amostra)
modelo_adequado_2 = step(modelo_zero_2, list(lower = ~ 1,
                                             upper = ~
x7+x10+log(x9)+x6+x4+x5+x8),
                        direction="forward")
```

```
## Start: AIC=-1898.38
## log(y) ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + log(x9)  1   125.213   24.299 -3713.3
## + x7       1    93.211   56.301 -2873.0
## + x8       1    68.513   80.999 -2509.3
## + x5       1    55.810   93.702 -2363.6
## + x6       1    54.978   94.534 -2354.8
## + x4       1    28.678  120.834 -2109.3
## + x10      1    17.764  131.748 -2022.9
## <none>             149.512 -1898.4
##
## Step: AIC=-3713.33
```



```

## log(y) ~ log(x9)
##
##      Df Sum of Sq  RSS    AIC
## + x8   1   10.9783 13.320 -4312.5
## + x10   1    4.6077 19.691 -3921.6
## + x4    1    4.4198 19.879 -3912.1
## + x5    1    4.1093 20.189 -3896.6
## + x6    1    3.8142 20.485 -3882.1
## + x7    1    3.6346 20.664 -3873.4
## <none>          24.299 -3713.3
##
## Step: AIC=-4312.45
## log(y) ~ log(x9) + x8
##
##      Df Sum of Sq  RSS    AIC
## + x7   1    5.1967  8.1238 -4805.0
## + x10   1    3.6793  9.6412 -4633.7
## + x4    1    1.3306 11.9899 -4415.7
## + x6    1    0.9316 12.3889 -4383.0
## + x5    1    0.8064 12.5141 -4372.9
## <none>          13.3205 -4312.5
##
## Step: AIC=-4804.96
## log(y) ~ log(x9) + x8 + x7
##
##      Df Sum of Sq  RSS    AIC
## + x10   1    3.07737 5.0464 -5279.1
## + x4    1    1.84001 6.2838 -5059.8
## + x6    1    1.25807 6.8657 -4971.2
## + x5    1    0.64829 7.4755 -4886.1
## <none>          8.1238 -4805.0
##
## Step: AIC=-5279.07
## log(y) ~ log(x9) + x8 + x7 + x10
##
##      Df Sum of Sq  RSS    AIC
## + x4    1    1.42444 3.6220 -5608.7
## + x6    1    0.96233 4.0841 -5488.7
## + x5    1    0.37509 4.6713 -5354.3
## <none>          5.0464 -5279.1
##
## Step: AIC=-5608.73
## log(y) ~ log(x9) + x8 + x7 + x10 + x4
##
##      Df Sum of Sq  RSS    AIC
## + x5    1    0.154421 3.4676 -5650.3
## <none>          3.6220 -5608.7
## + x6    1    0.000645 3.6214 -5606.9
##
## Step: AIC=-5650.3
## log(y) ~ log(x9) + x8 + x7 + x10 + x4 + x5
##
##      Df Sum of Sq  RSS    AIC

```

```
## <none>          3.4676 -5650.3
## + x6      1 0.00064646 3.4669 -5648.5

summary(modelo_adequado_2)

##
## Call:
## lm(formula = log(y) ~ log(x9) + x8 + x7 + x10 + x4 + x5, data =
## amostra)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11221 -0.02503 -0.00346  0.02302  0.52897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.3449384  0.0334706  -40.18  < 2e-16 ***
## log(x9)      1.6247313  0.0145829  111.41  < 2e-16 ***
## x8          -0.0295436  0.0005975  -49.45  < 2e-16 ***
## x7          -0.0311160  0.0008066  -38.58  < 2e-16 ***
## x10          0.1077881  0.0038545   27.96  < 2e-16 ***
## x4          -0.0053209  0.0002866  -18.57  < 2e-16 ***
## x5           0.0089942  0.0013525    6.65 4.83e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05909 on 993 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.9767
## F-statistic: 6970 on 6 and 993 DF, p-value: < 2.2e-16

plot(fitted(modelo_adequado_2), rstandard(modelo_adequado_2))
abline(0,0)
```

