Segunda lista de exercícios

Ernani de Souza Cubas Neto

12/06/2022

Com a base de dados “imoveiscwbav” obter os seguintes resultados com o auxílio do “R”

Carregando as bibliotecas

```
## Carregando pacotes exigidos: mgcv
```

```
## Carregando pacotes exigidos: nlme
```

```
## This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Carregando pacotes exigidos: sandwich
```

```
## Carregando pacotes exigidos: fit.models
```

```
## Registered S3 method overwritten by 'fit.models':
##   method       from
##   vcov.default Hmisc
```

---

a) Estimar um modelo preliminar e apresentar os resultados (“price”).

Estimando um modelo preliminar

```
##
## Call:
## lm(formula = formBase, data = imoveiscwbav)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -508060 -134595   -4235  105522 2414497
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -618334.1   156654.4  -3.947 9.00e-05 ***
## age           -7706.9     1032.1  -7.467 3.47e-13 ***
## parea          2566.9      628.8   4.082 5.16e-05 ***
## tarea          2016.0      335.8   6.004 3.62e-09 ***
## bath          16481.5    14926.2   1.104  0.27002
## ensuit       123180.4    18695.2   6.589 1.09e-10 ***
## garag        172322.6    21852.4   7.886 1.85e-14 ***
## plaz         141167.6    71440.9   1.976  0.04868 *
## park          -5577.8     4365.9  -1.278  0.20197
## trans          4858.8     3244.7   1.497  0.13488
## kidca           918.2     9386.8   0.098  0.92212
## school        21299.4    27893.7   0.764  0.44546
## health         2075.1    33517.9   0.062  0.95066
## bike         -59692.6    34069.4  -1.752  0.08035 .
## barb         -46467.8    22733.2  -2.044  0.04145 *
## balc          66863.0    25413.8   2.631  0.00877 **
## elev        -111708.3    25430.7  -4.393 1.36e-05 ***
## fitg         122356.6    28555.1   4.285 2.18e-05 ***
## party         35428.2    28633.7   1.237  0.21654
## categ        298785.7    56763.3   5.264 2.07e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 230900 on 521 degrees of freedom
## Multiple R-squared:  0.8073, Adjusted R-squared:  0.8003
## F-statistic: 114.9 on 19 and 521 DF,  p-value: < 2.2e-16
```

Sobre a resposta: Em um primeiro levantamento, verificamos que as variáveis “age”, “parea”, “tarea”, “ensuit”, “garag”, “elev”, “fitg” e “categ” possuem maiores graus de significância, ou seja, estatisticamente diferentes de zero. R^2 neste caso possui o valor de 0.8073, ou seja, as variáveis explicativas conseguem explicar 80,73% das variações da variável dependente (“price”). No caso do R^2 ajustado, o valor vai para 80,03% O valor da estatística F de 114.9 informa que pelo menos um dos parâmetros é diferente de 0, afirmando que existe uma reta de regressão

---

b) Testar as variáveis para formulação do modelo.

Testando as variáveis

```
## [1] "Testando a variável 'age':"
```

```
##                AIC      BIC ranking (BIC)
## sqr(x)    14887.65 14982.11           1.5
## x+x^2     14887.65 14982.11           1.5
## smoothing 14888.43 14986.09           3.0
## x^2       14901.13 14995.59           4.0
## log(x)    14909.85 15004.31           5.0
## base      14919.50 15009.66           6.5
## x         14919.50 15009.66           6.5
## 1/x       14921.50 15015.95           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included age in the base-formula and it is also the variable you
test"
```

```
## [1] "Testando a variável 'parea':"
```

```
##                AIC      BIC ranking (BIC)
## sqr(x)    14907.15 15001.60           1.5
## x+x^2     14907.15 15001.60           1.5
## x^2       14907.18 15001.63           3.0
## log(x)    14907.28 15001.73           4.0
## 1/x       14908.53 15002.99           5.0
## smoothing 14888.33 15008.88           6.0
## base      14919.50 15009.66           7.5
## x         14919.50 15009.66           7.5
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included parea in the base-formula and it is also the variable y
ou test"
```

```
## [1] "Testando a variável 'tarea':"
```

```
##                AIC      BIC ranking (BIC)
## sqr(x)    14868.11 14962.56           1.5
## x+x^2     14868.11 14962.56           1.5
## x^2       14875.20 14969.65           3.0
## smoothing 14870.71 14971.56           4.0
## log(x)    14878.54 14973.00           5.0
## 1/x       14893.50 14987.96           6.0
## base      14919.50 15009.66           7.5
## x         14919.50 15009.66           7.5
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included tarea in the base-formula and it is also the variable y
ou test"
```

```
## [1] "Testando a variável 'plaz':"
```

```
##                AIC      BIC ranking (BIC)
## base      14919.50 15009.66           1.5
## x         14919.50 15009.66           1.5
## sqr(x)    14919.68 15014.14           3.5
## x+x^2     14919.68 15014.14           3.5
## x^2       14919.89 15014.35           5.0
## log(x)    14919.95 15014.41           6.0
## 1/x       14920.13 15014.59           7.0
## smoothing 14912.09 15021.22           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included plaz in the base-formula and it is also the variable yo
u test"
```

```
## [1] "Testando a variável 'park':"
```

```
##                AIC      BIC ranking (BIC)
## 1/x       14899.04 14993.49           1.0
## log(x)    14905.43 14999.89           2.0
## x^2       14908.49 15002.95           3.0
## smoothing 14897.98 15004.01           4.0
## base      14919.50 15009.66           5.5
## x         14919.50 15009.66           5.5
## sqr(x)    14916.04 15010.49           7.5
## x+x^2     14916.04 15010.49           7.5
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included park in the base-formula and it is also the variable yo
u test"
```

```
## [1] "Testando a variável 'trans':"
```

```
##                AIC      BIC ranking (BIC)
## 1/x       14910.43 15004.88           1.0
## base      14919.50 15009.66           3.0
## x         14919.50 15009.66           3.0
## log(x)    14915.21 15009.66           3.0
## x^2       14916.90 15011.36           5.0
## sqr(x)    14920.12 15014.57           6.5
## x+x^2     14920.12 15014.57           6.5
## smoothing 14895.00 15018.12           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included trans in the base-formula and it is also the variable y
ou test"
```

```
## [1] "Testando a variável 'kidca':"
```

```
##                AIC      BIC ranking (BIC)
## sqr(x)    14907.48 15001.94           1.5
## x+x^2     14907.48 15001.94           1.5
## x^2       14910.97 15005.42           3.0
## log(x)    14912.64 15007.09           4.0
## base      14919.50 15009.66           5.5
## x         14919.50 15009.66           5.5
## 1/x       14918.72 15013.18           7.0
## smoothing 14894.67 15014.27           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included kidca in the base-formula and it is also the variable y
ou test"
```

```
## [1] "Testando a variável 'school':"
```

```
##                AIC      BIC ranking (BIC)
## base      14919.50 15009.66           1.5
## x         14919.50 15009.66           1.5
## sqr(x)    14921.36 15015.81           3.5
## x+x^2     14921.36 15015.81           3.5
## 1/x       14921.41 15015.87           5.0
## x^2       14921.44 15015.89           6.0
## log(x)    14921.44 15015.90           7.0
## smoothing 14918.82 15024.14           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included school in the base-formula and it is also the variable
you test"
```

```
## [1] "Testando a variável 'health':"
```

```
##                AIC      BIC ranking (BIC)
## base      14919.50 15009.66           1.5
## x         14919.50 15009.66           1.5
## smoothing 14919.36 15010.89           3.0
## x^2       14919.76 15014.21           4.0
## log(x)    14919.77 15014.22           5.0
## sqr(x)    14919.84 15014.29           6.5
## x+x^2     14919.84 15014.29           6.5
## 1/x       14919.97 15014.43           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006)
for an elaboration"
## [1] "please note that you included health in the base-formula and it is also the variable
you test"
```

```
## [1] "Testando a variável 'bike':"
```

```
##                AIC      BIC ranking (BIC)
## base      14919.50 15009.66           1.5
## x         14919.50 15009.66           1.5
## 1/x       14916.44 15010.90           3.0
## log(x)    14917.89 15012.35           4.0
## x^2       14918.27 15012.73           5.0
## sqr(x)    14919.34 15013.80           6.5
## x+x^2     14919.34 15013.80           6.5
## smoothing 14909.16 15031.81           8.0
## [1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
## [1] "please note that you included bike in the base-formula and it is also the variable you test"
```

c) Verifique a presença de outliers.

```
##      rstudent unadjusted p-value Bonferroni p
## 393 11.999157         1.8763e-29   1.0151e-26
## 13   4.412352         1.2437e-05   6.7287e-03
```

> **Sobre a resposta:** Foram detectadas as linhas 13 e 393 como outliers

d) Teste a especificação do modelo e altere se necessário.

```
##
##  RESET test
##
## data:  formBase
## RESET = 4.477, df1 = 38, df2 = 483, p-value = 2.103e-15
```

```
## [1] "F tabelado:"
```

```
## [1] 1.429987
```

> **Sobre a resposta:** O F Calculado é de 4.477, ou seja, maior que o F Tabelado. Nesse caso, ajustamos as variáveis de acordo com os testes em b)

Ajustando as variáveis

```
## [1] "sqrt(age)"
```

```
## [1] "sqrt(parea)"
```

```
## [1] "sqrt(tarea)"
```

```
## [1] "1/park"
```

```
## [1] "1/trans"
```

```
## [1] "sqrt(kidca)"
```

```
##
##  RESET test
##
## data:  formBase
## RESET = 4.2773, df1 = 38, df2 = 483, p-value = 1.942e-14
```

> **Sobre a resposta:** Mesmo realizando as transformações nas variáveis, conforme apresentado nos testes em b), o F calculado caiu para apenas 4.2773, continuando maior que o F tabelado. Tentaremos baixar este valor realizando os testes de multicolinearidade e exclusão de possíveis variáveis.

---

e) Teste a presença de multicolinearidade e exclua variáveis se necessário.

```
## [1] "Verificando pela matriz de correlação"
```

```
##                age       parea       tarea        bath      ensuit
## age     1.000000000 -0.23412397 -0.33529253 -0.25893809 -0.44034028
## parea  -0.234123966  1.00000000  0.83631214  0.68030305  0.59703635
## tarea  -0.335292527  0.83631214  1.00000000  0.65178059  0.56702436
## bath   -0.258938090  0.68030305  0.65178059  1.00000000  0.73616242
## ensuit -0.440340281  0.59703635  0.56702436  0.73616242  1.00000000
## garag  -0.458343501  0.60354894  0.64585494  0.57419774  0.53115159
## plaz    0.034885737 -0.05854644 -0.08806166 -0.05458843 -0.04102969
## park    0.049404976  0.22516497  0.23321585  0.18685239  0.04426073
## trans  -0.001292687 -0.26736118 -0.22145907 -0.17046710 -0.05379415
## kidca   0.040767689  0.23201885  0.25207028  0.22173506  0.14037251
## school  0.056237914  0.13302021  0.08045093  0.06006860  0.00640379
## health -0.066844568 -0.11986636 -0.12953939 -0.13899370 -0.11907669
## bike    0.068000907 -0.05858109 -0.04887671 -0.06131446 -0.05378576
##               garag         plaz        park        trans       kidca
## age    -0.458343501  0.034885737  0.04940498 -0.001292687  0.04076769
## parea   0.603548940 -0.058546435  0.22516497 -0.267361184  0.23201885
## tarea   0.645854936 -0.088061662  0.23321585 -0.221459070  0.25207028
## bath    0.574197743 -0.054588430  0.18685239 -0.170467100  0.22173506
## ensuit  0.531151588 -0.041029694  0.04426073 -0.053794150  0.14037251
## garag   1.000000000 -0.077867092  0.16444314 -0.144660588  0.12357575
## plaz   -0.077867092  1.000000000 -0.31214201  0.035948379 -0.06763570
## park    0.164443140 -0.312142011  1.00000000 -0.380701377  0.29905790
## trans  -0.144660588  0.035948379 -0.38070138  1.000000000 -0.50601287
## kidca   0.123575746 -0.067635695  0.29905790 -0.506012867  1.00000000
## school  0.006632926  0.142626552  0.44391529 -0.124481756  0.15454958
## health -0.024429728  0.009747807 -0.26075078  0.138609577 -0.34983769
## bike   -0.051137962  0.288391420 -0.35872405 -0.071974443  0.20178646
##              school       health        bike
## age     0.056237914 -0.066844568  0.06800091
## parea   0.133020211 -0.119866359 -0.05858109
## tarea   0.080450934 -0.129539385 -0.04887671
## bath    0.060068605 -0.138993705 -0.06131446
## ensuit  0.006403790 -0.119076688 -0.05378576
## garag   0.006632926 -0.024429728 -0.05113796
## plaz    0.142626552  0.009747807  0.28839142
## park    0.443915292 -0.260750781 -0.35872405
## trans  -0.124481756  0.138609577 -0.07197444
## kidca   0.154549581 -0.349837688  0.20178646
## school  1.000000000 -0.339536841  0.08137159
## health -0.339536841  1.000000000 -0.07429099
## bike    0.081371594 -0.074290985  1.00000000
```

```
## [1] "Excluindo 'tarea','bath','ensuit' e 'garag'"
```

```
## [1] "Verificando multicolinearidade pelo VIF - Valor de Inflação da Variância"
```

```
## VIFs computed for high-order terms
```

```
##      age    parea     plaz     park    trans    kidca   school   health
## 1.318456 1.347623 1.312498 2.430960 1.674674 1.717952 1.741332 1.292913
##     bike     barb     balc     elev     fitg    party    categ
## 1.560999 1.287502 1.611816 1.383136 1.697774 2.045691 1.274470
```

```
## [1] "Refazendo o RESETTest"
```

```
##
##  RESET test
##
## data:  formBase
## RESET = 2.9723, df1 = 30, df2 = 495, p-value = 4.732e-07
```

Sobre a resposta: Pela Matriz de Correlação, percebe-se correlações entre as variaveis ‘parea’com ’tarea’,‘bath’,‘ensuit’ e ‘garag’. Optou-se por retirar estas 4 últimas. Pelo teste VIF, já com as variáveis excluidas, não foram encontradas correlações. Refazendo o RESET test, o valor caiu para 2.9723

---

f) Selecione um modelo pela técnica de stepwise

```
##
## Call:
## lm(formula = formBase, data = imoveiscwbav)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -843239 -163941  -10706  120085 2574811
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1482116     217249  -6.822 2.48e-11 ***
## age          -110342       7600 -14.518  < 2e-16 ***
## parea         198699       9101  21.832  < 2e-16 ***
## plaz          190394      85150   2.236  0.02577 *
## park          640643     139463   4.594 5.46e-06 ***
## trans          -8264     126426  -0.065  0.94791
## kidca          17821      45459   0.392  0.69520
## school        -90172      34631  -2.604  0.00948 **
## health        -17586      38701  -0.454  0.64972
## bike           13725      42940   0.320  0.74938
## barb          -29053      26264  -1.106  0.26914
## balc           75723      29549   2.563  0.01067 *
## elev         -164242      29463  -5.575 3.97e-08 ***
## fitg          190178      32642   5.826 9.90e-09 ***
## party          66803      33241   2.010  0.04498 *
## categ         347120      63452   5.471 6.95e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 269200 on 525 degrees of freedom
## Multiple R-squared:  0.7363, Adjusted R-squared:  0.7287
## F-statistic:  97.7 on 15 and 525 DF,  p-value: < 2.2e-16
```

```
##
## Direction:  backward/forward
## Criterion:  AIC
##
## Start:  AIC=13544.13
## price ~ age + parea + plaz + park + trans + kidca + school +
##     health + bike + barb + balc + elev + fitg + party + categ
##
##          Df  Sum of Sq        RSS   AIC
## - trans   1 3.0957e+08 3.8039e+13 13542
## - bike    1 7.4022e+09 3.8046e+13 13542
## - kidca   1 1.1135e+10 3.8050e+13 13542
## - health  1 1.4961e+10 3.8054e+13 13542
## - barb    1 8.8664e+10 3.8127e+13 13543
## <none>                 3.8039e+13 13544
## - party   1 2.9263e+11 3.8331e+13 13546
## - plaz    1 3.6225e+11 3.8401e+13 13547
## - balc    1 4.7581e+11 3.8514e+13 13549
## - school  1 4.9121e+11 3.8530e+13 13549
## - park    1 1.5289e+12 3.9568e+13 13563
## - categ   1 2.1684e+12 4.0207e+13 13572
## - elev    1 2.2516e+12 4.0290e+13 13573
## - fitg    1 2.4594e+12 4.0498e+13 13576
## - age     1 1.5271e+13 5.3309e+13 13725
## - parea   1 3.4535e+13 7.2573e+13 13892
##
## Step:  AIC=13542.13
## price ~ age + parea + plaz + park + kidca + school + health +
##     bike + barb + balc + elev + fitg + party + categ
##
##          Df  Sum of Sq        RSS   AIC
## - bike    1 7.9248e+09 3.8047e+13 13540
## - health  1 1.4723e+10 3.8054e+13 13540
## - kidca   1 1.5149e+10 3.8054e+13 13540
## - barb    1 9.0491e+10 3.8129e+13 13541
## <none>                 3.8039e+13 13542
## + trans   1 3.0957e+08 3.8039e+13 13544
## - party   1 2.9232e+11 3.8331e+13 13544
## - plaz    1 3.6963e+11 3.8409e+13 13545
## - balc    1 4.7550e+11 3.8514e+13 13547
## - school  1 5.0543e+11 3.8544e+13 13547
## - park    1 1.7368e+12 3.9776e+13 13564
## - categ   1 2.1954e+12 4.0234e+13 13570
## - elev    1 2.2620e+12 4.0301e+13 13571
## - fitg    1 2.4906e+12 4.0530e+13 13574
## - age     1 1.5332e+13 5.3371e+13 13723
## - parea   1 3.5256e+13 7.3295e+13 13895
##
## Step:  AIC=13540.24
## price ~ age + parea + plaz + park + kidca + school + health +
##     barb + balc + elev + fitg + party + categ
##
##          Df  Sum of Sq        RSS   AIC
## - health  1 1.5327e+10 3.8062e+13 13538
## - kidca   1 2.6862e+10 3.8074e+13 13539
```

```
## - barb    1 8.4410e+10 3.8131e+13 13539
## <none>                 3.8047e+13 13540
## + bike    1 7.9248e+09 3.8039e+13 13542
## + trans   1 8.3218e+08 3.8046e+13 13542
## - party   1 2.9495e+11 3.8342e+13 13542
## - plaz    1 3.9101e+11 3.8438e+13 13544
## - balc    1 4.8176e+11 3.8529e+13 13545
## - school  1 5.0504e+11 3.8552e+13 13545
## - park    1 1.9987e+12 4.0046e+13 13566
## - categ   1 2.1900e+12 4.0237e+13 13568
## - elev    1 2.2550e+12 4.0302e+13 13569
## - fitg    1 2.4847e+12 4.0532e+13 13572
## - age     1 1.5422e+13 5.3469e+13 13722
## - parea   1 3.5381e+13 7.3428e+13 13894
##
## Step:  AIC=13538.46
## price ~ age + parea + plaz + park + kidca + school + barb + balc +
##     elev + fitg + party + categ
##
##          Df  Sum of Sq        RSS   AIC
## - kidca   1 4.3413e+10 3.8106e+13 13537
## - barb    1 8.5638e+10 3.8148e+13 13538
## <none>                 3.8062e+13 13538
## + health  1 1.5327e+10 3.8047e+13 13540
## + bike    1 8.5287e+09 3.8054e+13 13540
## + trans   1 4.0110e+08 3.8062e+13 13540
## - party   1 2.8745e+11 3.8350e+13 13540
## - plaz    1 3.8760e+11 3.8450e+13 13542
## - school  1 4.9367e+11 3.8556e+13 13543
## - balc    1 4.9612e+11 3.8558e+13 13544
## - park    1 2.0031e+12 4.0065e+13 13564
## - categ   1 2.2363e+12 4.0298e+13 13567
## - elev    1 2.2620e+12 4.0324e+13 13568
## - fitg    1 2.5022e+12 4.0564e+13 13571
## - age     1 1.5410e+13 5.3473e+13 13720
## - parea   1 3.5469e+13 7.3531e+13 13893
##
## Step:  AIC=13537.08
## price ~ age + parea + plaz + park + school + barb + balc + elev +
##     fitg + party + categ
##
##          Df  Sum of Sq        RSS   AIC
## - barb    1 8.2756e+10 3.8188e+13 13536
## <none>                 3.8106e+13 13537
## + kidca   1 4.3413e+10 3.8062e+13 13538
## + health  1 3.1878e+10 3.8074e+13 13539
## + bike    1 2.6260e+10 3.8079e+13 13539
## + trans   1 1.2755e+10 3.8093e+13 13539
## - party   1 2.9682e+11 3.8402e+13 13539
## - plaz    1 3.9215e+11 3.8498e+13 13541
## - balc    1 4.8056e+11 3.8586e+13 13542
## - school  1 4.8855e+11 3.8594e+13 13542
## - park    1 2.2016e+12 4.0307e+13 13566
## - elev    1 2.2547e+12 4.0360e+13 13566
## - categ   1 2.2924e+12 4.0398e+13 13567
## - fitg    1 2.5391e+12 4.0645e+13 13570
```

```
## - age     1 1.5374e+13 5.3480e+13 13718
## - parea   1 3.7353e+13 7.5458e+13 13905
##
## Step:  AIC=13536.25
## price ~ age + parea + plaz + park + school + balc + elev + fitg +
##     party + categ
##
##          Df  Sum of Sq        RSS   AIC
## <none>                3.8188e+13 13536
## + barb    1 8.2756e+10 3.8106e+13 13537
## + kidca   1 4.0532e+10 3.8148e+13 13538
## + health  1 3.2806e+10 3.8156e+13 13538
## + trans   1 1.6566e+10 3.8172e+13 13538
## + bike    1 1.3515e+10 3.8175e+13 13538
## - party   1 2.8463e+11 3.8473e+13 13538
## - balc    1 4.2115e+11 3.8609e+13 13540
## - plaz    1 4.3365e+11 3.8622e+13 13540
## - school  1 4.5815e+11 3.8646e+13 13541
## - categ   1 2.2812e+12 4.0470e+13 13566
## - elev    1 2.3163e+12 4.0505e+13 13566
## - park    1 2.3396e+12 4.0528e+13 13566
## - fitg    1 2.5197e+12 4.0708e+13 13569
## - age     1 1.5328e+13 5.3517e+13 13717
## - parea   1 3.7321e+13 7.5510e+13 13903
```

```
##
## Call:
## lm(formula = price ~ age + parea + plaz + park + school + balc +
##     elev + fitg + party + categ, data = imoveiscwbav)
##
## Coefficients:
## (Intercept)          age        parea         plaz         park       school
##    -1518007      -108534       199247       203965       658759       -80392
##        balc         elev         fitg        party        categ
##       69441      -165569       189942        65580       349594
```

Sobre a resposta: Segundo a técnica de stepwise, o modelo com o menor AIC,e portanto o melhor modelo considerado, é o price ~ age + parea + plaz + park + school + balc + elev + fitg + party + categ.

Retestando o novo modelo

```
##
## Call:
## lm(formula = formBase, data = imoveiscwbav)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -810132 -157741  -15850  116309 2575101
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1518007     161341  -9.409  < 2e-16 ***
## age          -108534       7441 -14.586  < 2e-16 ***
## parea         199247       8755  22.759  < 2e-16 ***
## plaz          203965      83141   2.453   0.0145 *
## park          658759     115606   5.698 2.01e-08 ***
## school        -80392      31881  -2.522   0.0120 *
## balc           69441      28723   2.418   0.0160 *
## elev         -165569      29202  -5.670 2.35e-08 ***
## fitg          189942      32120   5.913 6.01e-09 ***
## party          65580      32996   1.988   0.0474 *
## categ         349594      62132   5.627 2.98e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 268400 on 530 degrees of freedom
## Multiple R-squared:  0.7352, Adjusted R-squared:  0.7302
## F-statistic: 147.2 on 10 and 530 DF,  p-value: < 2.2e-16
```

```
##
##  RESET test
##
## data:  formBase
## RESET = 2.8348, df1 = 20, df2 = 510, p-value = 4.555e-05
```

> **Sobre a resposta:** Após as alterações no modelo definidas pelo stepwise, o F caiu para 2.8348

---

g) Faça o teste de homocedasticidade e faça correção da heterocedasticidade se necessário

```
##
##  Breusch-Pagan test
##
## data:  formBase
## BP = 197.73, df = 10, p-value < 2.2e-16
```

```
## [1] "O valor chiquadrado tabelado é:"
```

```
## [1] 18.30704
```

Sobre a resposta: o resultado do teste BP foi de 197.73. Por ser maior que o valor chiquadrado tabelado, rejeita-se a hipótese de homocedasticidade

Reduzindo a variância das variáveis por meio do log das variáveis:

```
##
##  Breusch-Pagan test
##
## data:  formBaseL
## BP = 35.143, df = 10, p-value = 0.000118
```

Sobre a resposta: Reduzindo a variância das variáveis, o valor em BP caiu. Porém, continua maior que o valor chiquadrado tabelado.

Regressão normal (com a redução das variância das variáveis)

```
##
## Call:
## lm(formula = formBaseL, data = imoveiscwbav)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81947 -0.14520  0.00767  0.13929  0.93967
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.91716    0.21190  37.362  < 2e-16 ***
## lage        -0.33246    0.01695 -19.611  < 2e-16 ***
## lparea       2.47613    0.07521  32.923  < 2e-16 ***
## lplaz        0.14744    0.08715   1.692  0.09126 .
## lpark        0.12058    0.02037   5.918 5.85e-09 ***
## lschool     -0.15101    0.04818  -3.134  0.00182 **
## balc         0.07510    0.02306   3.256  0.00120 **
## elev        -0.12145    0.02354  -5.159 3.51e-07 ***
## fitg         0.13470    0.02581   5.219 2.58e-07 ***
## party        0.10553    0.02659   3.969 8.22e-05 ***
## categ        0.34643    0.05028   6.890 1.59e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2159 on 530 degrees of freedom
## Multiple R-squared:  0.8334, Adjusted R-squared:  0.8302
## F-statistic: 265.1 on 10 and 530 DF,  p-value: < 2.2e-16
```

Regressão robusta

```
##
## Call:
## lmRob(formula = formBaseL, data = imoveiscwbav)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.844810 -0.140889  0.003637  0.131441  0.942078
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.71584    0.21020  36.707  < 2e-16 ***
## lage        -0.34177    0.01680 -20.343  < 2e-16 ***
## lparea       2.53060    0.07429  34.063  < 2e-16 ***
## lplaz        0.21014    0.08588   2.447 0.014733 *
## lpark        0.12565    0.01998   6.290 6.65e-10 ***
## lschool     -0.16369    0.04713  -3.473 0.000557 ***
## balc         0.06916    0.02260   3.060 0.002327 **
## elev        -0.10084    0.02306  -4.372 1.48e-05 ***
## fitg         0.12006    0.02519   4.767 2.42e-06 ***
## party        0.08793    0.02601   3.381 0.000776 ***
## categ        0.44864    0.05187   8.650  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1972 on 530 degrees of freedom
## Multiple R-Squared: 0.6908
##
## Test for Bias:
##             statistic   p-value
## M-estimate      19.37 0.0547545
## LS-estimate     35.07 0.0002411
```

h) Obtenha os indicadores de desempenho do modelo

```
## # Indices of model performance
##
## AIC      |     BIC |    R2 | R2 (adj.) |  RMSE | Sigma
## ----------------------------------------------------------
## -110.667 | -59.146 | 0.833 |     0.830 | 0.214 | 0.216
```

```
## # Indices of model performance
##
## R2    |  RMSE | Sigma
## ---------------------
## 0.831 | 0.215 | 0.217
```

Sobre a resposta: Para o modelo de regressão normal, tem-se os valores de RMSE e Sigma ligeiramente menores, se comparado os mesmos indicadores para o modelo robusto. No caso do R^2, o valor é 0,002 maior.

i) Estime os intervalos de confiança para os parâmetros do modelo

Intevalos de confiança - Regressão Normal

```
##                   2.5 %      97.5 %
## (Intercept)  7.50088861  8.33343110
## lage        -0.36576005 -0.29915598
## lparea       2.32838249  2.62387469
## lplaz       -0.02375661  0.31864076
## lpark        0.08055638  0.16060574
## lschool     -0.24566273 -0.05635416
## balc         0.02979603  0.12040115
## elev        -0.16769800 -0.07520800
## fitg         0.08399803  0.18539553
## party        0.05329568  0.15776308
## categ        0.24765330  0.44519702
```

Intevalos de confiança - Regressão Robusta

```
##                   2.5 %      97.5 %
## (Intercept)  7.30385141  8.12783001
## lage        -0.37470046 -0.30884279
## lparea       2.38499196  2.67620645
## lplaz        0.04181829  0.37846772
## lpark        0.08650006  0.16480629
## lschool     -0.25606779 -0.07130245
## balc         0.02486014  0.11346435
## elev        -0.14603913 -0.05563204
## fitg         0.07069698  0.16942962
## party        0.03695769  0.13890751
## categ        0.34698348  0.55029257
```

---

j) Faça predição de um imóvel hipotético: apresente seus parâmetros de simulação e o resultado

Para o teste, foram selecionados os valores das medianas das variáveis da base (para a predição é necessário substituir o valor pelo seu log, conforme conversão realizada anteriormente):

- age = 3 -> log(3) = 1.0986
- parea = 10.954 -> log(10.954) = 2.394
- plaz = 0.20671 -> log(0.20671) = -1.5764
- park = 3.5507 - > log(3.5507) = 1.2672
- school = 1.764 -> log(1.764) = 0.5675
- balc = 0
- elev = 0
- fitg = 0
- party = 1
- categ = 1

A mediana de ‘price’ é 880000, ou log(880000) = 13.69

```
val <- predict(object = resultrob,
         data.frame(lage=1.0986, lparea=2.394,lplaz=-1.5764,lpark=1.2672,lschool=0.5675,balc=
0,elev=0,fitg=0,party=1,categ=1))
val
```

```
##        1
## 13.67026
```

Convertendo o valor

```
exp(13.67026)
```

```
## [1] 864805.6
```