Estatística 1

Segunda lista de exercício

a) Estimar um modelo preliminar e apresentar os resultados;

```
> load("C:/Users/bruno/Documents/Facul/lista2/imoveiscwbav.RData")
> gc()
         used (Mb) gc trigger (Mb) max used (Mb)
Ncells 499840 26.7    1108625 59.3   638977 34.2
Vcells 778866  6.0    8388608 64.0  1631704 12.5
> resultados <- lm(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ,data=imoveiscwbav)
>
>
> summary (resultados)

Call:
lm(formula = price ~ age + parea + tarea + bath + ensuit + garag +
    plaz + park + trans + kidca + school + health + bike + barb +
    balc + elev + fitg + party + categ, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-495718 -134211   -2632  104528 2419265

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -420453.5   130052.5  -3.233   0.0013 **
age           -7839.1     1025.3  -7.645 1.01e-13 ***
parea          2592.2      624.0   4.154 3.82e-05 ***
tarea          1975.8      333.9   5.918 5.91e-09 ***
bath          13452.6    14832.9   0.907   0.3649
ensuit       125949.6    18560.7   6.786 3.15e-11 ***
garag        169687.5    21756.1   7.800 3.41e-14 ***
plaz         224393.0    94219.1   2.382   0.0176 *
park         -63439.6    27154.0  -2.336   0.0199 *
trans         26642.3    22718.5   1.173   0.2414
kidca         10452.8    34899.8   0.300   0.7647
school        -7975.8    56635.7  -0.141   0.8881
health         1217.4    56216.5   0.022   0.9827
bike         -85864.4    56073.0  -1.531   0.1263
barb         -43925.7    22602.3  -1.943   0.0525 .
balc          65144.8    25242.3   2.581   0.0101 *
elev        -111743.4    25295.0  -4.418 1.21e-05 ***
fitg         123052.7    28456.0   4.324 1.83e-05 ***
party         36463.1    28481.1   1.280   0.2010
categ        283061.5    55653.0   5.086 5.11e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229400 on 521 degrees of freedom
Multiple R-squared:  0.8099,    Adjusted R-squared:  0.803
F-statistic: 116.8 on 19 and 521 DF,  p-value: < 2.2e-16
```

Carreguei a base e criei o modelo.

b) Testar as variáveis para formulação do modelo;

Teste Age

```
>
> formBase<-formula(price~parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-559997 -135850   -7282  108797 2360349

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -593909.9   134921.3  -4.402 1.30e-05 ***
parea          1643.3      644.3   2.550 0.011042 *
tarea          2396.9      346.9   6.909 1.43e-11 ***
bath          -8495.5    15332.4  -0.554 0.579756
ensuit       172325.9    18481.5   9.324  < 2e-16 ***
garag        215693.1    22027.9   9.792  < 2e-16 ***
plaz         250882.7    99201.5   2.529 0.011732 *
park         -51641.8    28563.1  -1.808 0.071183 .
trans         36393.6    23898.3   1.523 0.128401
kidca          4784.5    36762.0   0.130 0.896500
school        -2905.4    59667.0  -0.049 0.961182
health        35316.9    59042.6   0.598 0.549993
bike        -132282.7    58730.9  -2.252 0.024715 *
barb         -21661.4    23615.2  -0.917 0.359427
balc          91190.0    26351.8   3.460 0.000583 ***
elev         -91613.1    26505.9  -3.456 0.000592 ***
fitg         140020.5    29889.7   4.685 3.58e-06 ***
party         34036.1    30005.6   1.134 0.257178
categ        225297.9    58092.8   3.878 0.000119 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 241700 on 522 degrees of freedom
Multiple R-squared:  0.7886,    Adjusted R-squared:  0.7813
F-statistic: 108.2 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenage<-fform(imoveiscwbav,"age",formBase)
               AIC      BIC ranking (BIC)
smoothing 14879.74 14977.41             1
x^2       14897.14 14987.30             2
log(x)    14901.06 14991.22             3
x         14912.25 15002.42             4
x+x^2     14943.68 15033.85             5
sqr(x)    14944.35 15034.51             6
1/x       14957.80 15047.96             7
base      14967.78 15053.65             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Incluir Age

```
> formBase<-formula(price~age+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-536220 -128848    1494  108580 2420400

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -345157.5   130773.1  -2.639  0.00855 **
age           -6991.9     1020.4  -6.852 2.06e-11 ***
tarea          2853.5      262.5  10.870  < 2e-16 ***
bath          23570.7    14857.7   1.586  0.11325
ensuit       139534.2    18552.6   7.521 2.39e-13 ***
garag        180717.9    21927.1   8.242 1.38e-15 ***
plaz         225240.4    95674.8   2.354  0.01893 *
park         -56223.2    27517.1  -2.043  0.04153 *
trans         41171.8    22794.5   1.806  0.07146 .
kidca          9002.4    35437.3   0.254  0.79957
school        23779.1    56984.6   0.417  0.67664
health         3818.0    57081.6   0.067  0.94670
bike        -105147.8    56744.0  -1.853  0.06444 .
barb         -41841.0    22945.9  -1.823  0.06881 .
balc          67611.1    25625.3   2.638  0.00858 **
elev        -110314.7    25683.5  -4.295 2.08e-05 ***
fitg         116193.4    28847.0   4.028 6.46e-05 ***
party         35352.9    28919.9   1.222  0.22209
categ        225210.7    54715.0   4.116 4.48e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 232900 on 522 degrees of freedom
Multiple R-squared:  0.8036,    Adjusted R-squared:  0.7968
F-statistic: 118.7 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenparea<-fform(imoveiscwbav,"parea",formBase)
               AIC       BIC ranking (BIC)
sqr(x)    14904.77 14994.94             1
x+x^2     14904.80 14994.96             2
x         14912.25 15002.42             3
smoothing 14881.96 15002.43             4
x^2       14916.70 15006.87             5
log(x)    14921.03 15011.19             6
base      14927.88 15013.75             7
1/x       14929.78 15019.94             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
>
```

Teste Parea

```
> formBase<-formula(price~age+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-536220 -128848    1494  108580 2420400

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -345157.5   130773.1  -2.639  0.00855 **
age           -6991.9     1020.4  -6.852 2.06e-11 ***
tarea          2853.5      262.5  10.870  < 2e-16 ***
bath          23570.7    14857.7   1.586  0.11325
ensuit       139534.2    18552.6   7.521 2.39e-13 ***
garag        180717.9    21927.1   8.242 1.38e-15 ***
plaz         225240.4    95674.8   2.354  0.01893 *
park         -56223.2    27517.1  -2.043  0.04153 *
trans         41171.8    22794.5   1.806  0.07146 .
kidca          9002.4    35437.3   0.254  0.79957
school        23779.1    56984.6   0.417  0.67664
health         3818.0    57081.6   0.067  0.94670
bike        -105147.8    56744.0  -1.853  0.06444 .
barb         -41841.0    22945.9  -1.823  0.06881 .
balc          67611.1    25625.3   2.638  0.00858 **
elev        -110314.7    25683.5  -4.295 2.08e-05 ***
fitg         116193.4    28847.0   4.028 6.46e-05 ***
party         35352.9    28919.9   1.222  0.22209
categ        225210.7    54715.0   4.116 4.48e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 232900 on 522 degrees of freedom
Multiple R-squared:  0.8036,    Adjusted R-squared:  0.7968
F-statistic: 118.7 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenparea<-fform(imoveiscwbav,"parea",formBase)
               AIC      BIC ranking (BIC)
sqr(x)    14904.77 14994.94             1
x+x^2     14904.80 14994.96             2
x         14912.25 15002.42             3
smoothing 14881.96 15002.43             4
x^2       14916.70 15006.87             5
log(x)    14921.03 15011.19             6
base      14927.88 15013.75             7
1/x       14929.78 15019.94             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
>
```

Decisão: Incluir Parea

Teste Tarea

```
> formBase<-formula(price~age+parea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-473778 -125283   -4964   95015 2459849

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -425254.9   134221.1  -3.168  0.00162 **
age           -8840.2     1043.7  -8.470 2.51e-16 ***
parea          4929.2      498.7   9.885  < 2e-16 ***
bath          22083.6    15234.5   1.450  0.14778
ensuit       122215.9    19144.9   6.384 3.82e-10 ***
garag        195599.8    21994.5   8.893  < 2e-16 ***
plaz         215182.2    97227.8   2.213  0.02732 *
park         -75332.0    27948.1  -2.695  0.00726 **
trans         19109.4    23410.3   0.816  0.41471
kidca         26619.7    35908.7   0.741  0.45884
school       -24685.9    58379.6  -0.423  0.67258
health        -7441.8    57999.9  -0.128  0.89796
bike         -74259.7    57836.1  -1.284  0.19972
barb         -44110.9    23327.2  -1.891  0.05918 .
balc          62890.8    26049.0   2.414  0.01611 *
elev        -119874.1    26067.8  -4.599 5.34e-06 ***
fitg         122237.3    29368.3   4.162 3.69e-05 ***
party         45371.6    29353.5   1.546  0.12278
categ        339852.4    56577.6   6.007 3.55e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 236800 on 522 degrees of freedom
Multiple R-squared:  0.7971,    Adjusted R-squared:  0.7901
F-statistic: 113.9 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJentarea<-fform(imoveiscwbav,"tarea",formBase)
               AIC      BIC ranking (BIC)
smoothing 14863.46 14963.98             1
sqr(x)    14886.91 14977.07             2
x+x^2     14886.96 14977.12             3
x         14912.25 15002.42             4
x^2       14925.17 15015.33             5
log(x)    14935.80 15025.96             6
base      14945.45 15031.32             7
1/x       14947.12 15037.28             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Incluir Tarea

Teste Plaz

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+park+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-487535 -137496   -4432  108110 2436449

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -445551.8   130203.6  -3.422  0.00067 ***
age           -7928.9     1029.2  -7.704 6.69e-14 ***
parea          2595.4      626.8   4.141 4.04e-05 ***
tarea          1962.7      335.3   5.853 8.53e-09 ***
bath          14915.5    14886.4   1.002  0.31683
ensuit       123964.4    18624.7   6.656 7.14e-11 ***
garag        170152.7    21852.4   7.786 3.73e-14 ***
park         -42810.2    25850.3  -1.656  0.09831 .
trans         29699.8    22783.5   1.304  0.19296
kidca          5371.1    34990.1   0.154  0.87806
school        32969.5    54204.2   0.608  0.54329
health         4238.3    56453.1   0.075  0.94018
bike         -67380.9    55781.3  -1.208  0.22761
barb         -49106.4    22597.8  -2.173  0.03022 *
balc          67618.0    25333.6   2.669  0.00784 **
elev        -116076.3    25342.2  -4.580 5.81e-06 ***
fitg         121377.1    28574.3   4.248 2.56e-05 ***
party         39742.0    28574.8   1.391  0.16488
categ        294370.6    55697.6   5.285 1.85e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 230400 on 522 degrees of freedom
Multiple R-squared:  0.8078,    Adjusted R-squared:  0.8012
F-statistic: 121.9 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenplaz<-fform(imoveiscwbav,"plaz",formBase)
              AIC      BIC ranking (BIC)
log(x)    14910.84 15001.00             1
x^2       14911.34 15001.50             2
base      14916.11 15001.98             3
x         14912.25 15002.42             4
x+x^2     14912.90 15003.06             5
sqr(x)    14914.25 15004.41             6
1/x       14915.88 15006.04             7
smoothing 14904.11 15015.66             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Não incluir Plaz

Teste Park

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+
+                       trans+kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-485013 -143277  -12614  107799 2401659

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -619461.9    98692.8  -6.277 7.28e-10 ***
age           -7703.0     1028.1  -7.493 2.90e-13 ***
parea          2498.9      625.4   3.996 7.37e-05 ***
tarea          2033.5      334.4   6.082 2.30e-09 ***
bath          15777.2    14862.6   1.062  0.28894
ensuit       121628.3    18547.0   6.558 1.31e-10 ***
garag        173458.1    21788.6   7.961 1.07e-14 ***
plaz         154175.0    89677.3   1.719  0.08617 .
trans         52060.7    20028.9   2.599  0.00961 **
kidca         11610.9    35045.0   0.331  0.74054
school        55960.3    49796.4   1.124  0.26162
health       -10386.0    56235.3  -0.185  0.85354
bike        -133014.7    52538.1  -2.532  0.01164 *
barb         -42641.7    22691.9  -1.879  0.06078 .
balc          70770.2    25234.3   2.805  0.00523 **
elev        -112024.2    25402.5  -4.410 1.26e-05 ***
fitg         121885.1    28572.8   4.266 2.37e-05 ***
party         36010.5    28601.8   1.259  0.20858
categ        313021.2    54386.2   5.756 1.47e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 230400 on 522 degrees of freedom
Multiple R-squared:  0.8079,    Adjusted R-squared:  0.8013
F-statistic:   122 on 18 and 522 DF,  p-value: < 2.2e-16

> PanJenpark<-fform(imoveiscwbav,"park",formBase)
                AIC      BIC ranking (BIC)
1/x       14905.25 14995.41             1
smoothing 14895.01 14995.66             2
log(x)    14906.77 14996.93             3
x^2       14909.76 14999.92             4
base      14915.89 15001.76             5
x         14912.25 15002.42             6
x+x^2     14914.73 15004.89             7
sqr(x)    14915.32 15005.48             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Incluir Park

Teste Trans

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       kidca+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-499136 -132980   -5716  104035 2418690

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -357200.0   118382.6  -3.017  0.00267 **
age           -7906.6     1024.1  -7.721 5.95e-14 ***
parea          2704.9      616.8   4.385 1.40e-05 ***
tarea          1953.9      333.5   5.859 8.24e-09 ***
bath          13235.2    14837.1   0.892  0.37279
ensuit       124875.6    18544.7   6.734 4.37e-11 ***
garag        170325.1    21757.1   7.828 2.77e-14 ***
plaz         230636.8    94102.4   2.451  0.01458 *
park         -78689.5    23846.3  -3.300  0.00103 **
kidca         25866.4    32341.9   0.800  0.42420
school       -21418.6    55483.5  -0.386  0.69963
health        10519.7    55674.0   0.189  0.85020
bike         -77781.5    55667.8  -1.397  0.16293
barb         -46944.8    22463.3  -2.090  0.03711 *
balc          64635.0    25247.6   2.560  0.01075 *
elev        -109777.7    25248.5  -4.348 1.65e-05 ***
fitg         119703.0    28322.4   4.226 2.80e-05 ***
party         35384.5    28476.4   1.243  0.21458
categ        271600.5    54807.8   4.956 9.76e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229500 on 522 degrees of freedom
Multiple R-squared:  0.8094,    Adjusted R-squared:  0.8028
F-statistic: 123.2 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJentrans<-fform(imoveiscwbav,"trans",formBase)
               AIC      BIC ranking (BIC)
1/x       14906.86 14997.02             1
base      14911.68 14997.55             2
log(x)    14910.50 15000.66             3
x^2       14911.49 15001.65             4
smoothing 14901.64 15001.97             5
x         14912.25 15002.42             6
x+x^2     14912.98 15003.14             7
sqr(x)    14913.14 15003.31             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Não incluir Trans

Teste Kidca

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+school+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-502179 -133678   -1484  105125 2419604

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -414146.4   128224.3  -3.230  0.00132 **
age           -7832.6     1024.2  -7.647 9.94e-14 ***
parea          2590.3      623.4   4.155 3.80e-05 ***
tarea          1983.6      332.6   5.965 4.52e-09 ***
bath          13727.7    14791.5   0.928  0.35380
ensuit       126023.1    18542.8   6.796 2.94e-11 ***
garag        169361.9    21710.0   7.801 3.36e-14 ***
plaz         222667.7    93960.8   2.370  0.01816 *
park         -63555.1    27127.6  -2.343  0.01951 *
trans         29204.9    21027.5   1.389  0.16546
school        -8130.1    56584.0  -0.144  0.88581
health        -3562.0    53857.1  -0.066  0.94729
bike         -82185.7    54663.6  -1.503  0.13332
barb         -43846.3    22581.0  -1.942  0.05271 .
balc          64650.0    25166.2   2.569  0.01048 *
elev        -111936.0    25264.8  -4.431 1.15e-05 ***
fitg         123800.3    28321.5   4.371 1.49e-05 ***
party         36872.0    28423.5   1.297  0.19512
categ        285093.1    55189.8   5.166 3.41e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229200 on 522 degrees of freedom
Multiple R-squared:  0.8099,    Adjusted R-squared:  0.8033
F-statistic: 123.5 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenkidca<-fform(imoveiscwbav,"kidca",formBase)
               AIC      BIC ranking (BIC)
base      14910.35 14996.22             1
log(x)    14912.07 15002.23             2
x^2       14912.13 15002.29             3
x         14912.25 15002.42             4
sqr(x)    14912.32 15002.48             5
1/x       14912.34 15002.50             6
x+x^2     14912.35 15002.51             7
smoothing 14887.98 15005.85             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Não incluir Kidca

Teste School

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+health+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-496324 -134050   -1325  103641 2419050

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -430263.7   109719.3  -3.921 9.97e-05 ***
age           -7837.4     1024.3  -7.651 9.66e-14 ***
parea          2580.4      617.7   4.177 3.46e-05 ***
tarea          1978.2      333.1   5.938 5.27e-09 ***
bath          13506.2    14814.1   0.912  0.36234
ensuit       126058.6    18527.1   6.804 2.80e-11 ***
garag        169846.7    21706.3   7.825 2.84e-14 ***
plaz         220365.2    89688.8   2.457  0.01433 *
park         -61591.8    23751.2  -2.593  0.00978 **
trans         27289.8    22227.4   1.228  0.22009
kidca         10497.5    34865.6   0.301  0.76347
health         3229.0    54320.3   0.059  0.95262
bike         -87242.5    55160.7  -1.582  0.11435
barb         -43472.0    22350.6  -1.945  0.05231 .
balc          65403.6    25151.7   2.600  0.00958 **
elev        -111690.8    25268.5  -4.420 1.20e-05 ***
fitg         122548.0    28202.9   4.345 1.67e-05 ***
party         36431.2    28453.4   1.280  0.20098
categ        284684.4    54395.5   5.234 2.41e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229200 on 522 degrees of freedom
Multiple R-squared:  0.8099,    Adjusted R-squared:  0.8033
F-statistic: 123.6 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenschool<-fform(imoveiscwbav,"school",formBase)
               AIC      BIC ranking (BIC)
base      14910.28 14996.14             1
1/x       14908.59 14998.75             2
log(x)    14912.09 15002.26             3
x^2       14912.21 15002.38             4
x         14912.25 15002.42             5
x+x^2     14912.27 15002.43             6
sqr(x)    14912.27 15002.44             7
smoothing 14902.71 15024.09             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Não incluir School

Teste Heath

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+bike+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-495906 -134106   -2849  104632 2419065

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -419586.4   123616.3  -3.394  0.00074 ***
age           -7840.9     1021.1  -7.679 7.99e-14 ***
parea          2592.4      623.4   4.159 3.74e-05 ***
tarea          1975.6      333.4   5.925 5.67e-09 ***
bath          13448.6    14817.6   0.908  0.36450
ensuit       125913.1    18466.1   6.819 2.55e-11 ***
garag        169707.3    21716.0   7.815 3.05e-14 ***
plaz         224439.0    94104.9   2.385  0.01744 *
park         -63387.7    27021.9  -2.346  0.01936 *
trans         26711.7    22469.7   1.189  0.23506
kidca         10238.3    33432.2   0.306  0.75954
school        -8287.5    54724.4  -0.151  0.87969
bike          -85972.2    55798.1  -1.541  0.12398
barb          -43910.8    22570.1  -1.946  0.05225 .
balc          65104.2    25148.2   2.589  0.00990 **
elev        -111748.8    25269.6  -4.422 1.19e-05 ***
fitg         123050.8    28428.6   4.328 1.80e-05 ***
party         36511.1    28367.5   1.287  0.19864
categ        283006.2    55541.1   5.095 4.87e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229200 on 522 degrees of freedom
Multiple R-squared:  0.8099,    Adjusted R-squared:  0.8034
F-statistic: 123.6 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenhealth<-fform(imoveiscwbav,"health",formBase)
              AIC      BIC ranking (BIC)
base      14910.26 14996.12           1.0
1/x       14911.26 15001.42           2.0
sqr(x)    14912.08 15002.24           3.0
x+x^2     14912.21 15002.37           4.5
log(x)    14912.21 15002.37           4.5
x^2       14912.23 15002.39           6.0
x         14912.25 15002.42           7.0
smoothing 14911.60 15004.47           8.0
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
```

Decisão: Não incluir Heath

Teste Bike

```
> formBase<-formula(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                       trans+kidca+school+health+barb+balc+elev+
+                       fitg+party+categ)
>
> summary(lm(formBase, data=imoveiscwbav))

Call:
lm(formula = formBase, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-510703 -134525      33  104312 2420136

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -399412.6   129491.1  -3.084  0.00215 **
age           -8009.1     1020.6  -7.847 2.42e-14 ***
parea          2671.3      622.7   4.290 2.13e-05 ***
tarea          1957.9      334.1   5.860 8.19e-09 ***
bath          14345.4    14840.5   0.967  0.33418
ensuit       126035.0    18584.5   6.782 3.22e-11 ***
garag        168530.2    21771.0   7.741 5.15e-14 ***
plaz         204424.0    93432.4   2.188  0.02912 *
park         -78405.3    25366.9  -3.091  0.00210 **
trans         22366.0    22575.3   0.991  0.32228
kidca         -1253.1    34096.2  -0.037  0.97070
school       -23111.0    55838.4  -0.414  0.67912
health         8860.3    56066.6   0.158  0.87449
barb         -50094.9    22269.0  -2.250  0.02489 *
balc          64100.9    25265.6   2.537  0.01147 *
elev        -113025.0    25313.7  -4.465 9.82e-06 ***
fitg         125879.2    28432.6   4.427 1.16e-05 ***
party         34485.7    28488.4   1.211  0.22663
categ        288705.2    55602.3   5.192 2.98e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 229700 on 522 degrees of freedom
Multiple R-squared:  0.8091,    Adjusted R-squared:  0.8025
F-statistic: 122.9 on 18 and 522 DF,  p-value: < 2.2e-16

>
> PanJenbike<-fform(imoveiscwbav,"bike",formBase)
               AIC      BIC ranking (BIC)
base      14912.68 14998.55             1
log(x)    14910.52 15000.68             2
x^2       14911.06 15001.22             3
x         14912.25 15002.42             4
x+x^2     14912.97 15003.13             5
sqr(x)    14913.65 15003.82             6
1/x       14914.12 15004.28             7
smoothing 14909.28 15007.97             8
[1] "Smoothing is a semi-parametric and data-driven transformation, please see Wood (2006) for an elaboration"
>
```

Decisão: Não incluir Bike

Resultado

```
> imoveiscwbav$age <- (imoveiscwbav$age^2)
> imoveiscwbav$parea <- sqrt(imoveiscwbav$parea)
> imoveiscwbav$tarea <- sqrt(imoveiscwbav$tarea)
> imoveiscwbav$plaz <- log(imoveiscwbav$plaz)
> imoveiscwbav$park <- 1/(imoveiscwbav$park)
> imoveiscwbav$trans <- 1/(imoveiscwbav$trans)
>
>
> save(imoveiscwbav, file="imoveiscwbav1.RData" )
> resultados <- lm(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+                      trans+kidca+school+health+bike+barb+balc+elev+
+                      fitg+party+categ,data=imoveiscwbav)
>
>
> summary (resultados)

Call:
lm(formula = price ~ age + parea + tarea + bath + ensuit + garag +
    plaz + park + trans + kidca + school + health + bike + barb +
    balc + elev + fitg + party + categ, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-575422 -132235   -1347   98967 2408789

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1011256.2   141686.1  -7.137 3.21e-12 ***
age              -129.3       25.3  -5.110 4.54e-07 ***
parea          46751.5    14519.8   3.220  0.00136 **
tarea          55178.0     9758.0   5.655 2.58e-08 ***
bath            4079.9    15284.1   0.267  0.78962
ensuit        153095.5    18852.5   8.121 3.37e-15 ***
garag         184403.4    22478.4   8.204 1.84e-15 ***
plaz           53988.2    18080.8   2.986  0.00296 **
park          190747.3    50923.4   3.746  0.00020 ***
trans         -89961.3    44710.8  -2.012  0.04473 *
kidca         -27557.6    37516.0  -0.735  0.46294
school        -67472.8    62015.7  -1.088  0.27710
health         16456.8    57633.3   0.286  0.77534
bike          -45288.4    61323.6  -0.739  0.46053
barb          -36985.5    23387.1  -1.581  0.11438
balc           79093.6    25959.3   3.047  0.00243 **
elev         -106901.9    26176.1  -4.084 5.12e-05 ***
fitg          139061.0    29377.3   4.734 2.85e-06 ***
party          23647.6    29530.0   0.801  0.42361
categ         259957.1    56205.5   4.625 4.73e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 237200 on 521 degrees of freedom
Multiple R-squared:  0.7967,    Adjusted R-squared:  0.7893
F-statistic: 107.5 on 19 and 521 DF,  p-value: < 2.2e-16
```

c) Verifique a presença de outliers;

```
> outlierTest(resultados)
     rstudent unadjusted p-value Bonferroni p
393 11.549391         1.2643e-27   6.8401e-25
13   4.413506         1.2374e-05   6.6942e-03
```

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

```
> resettest(price~age+parea+tarea+bath+ensuit+garag+plaz+park+
+              trans+kidca+school+health+bike+barb+balc+elev+
+              fitg+party+categ,power=2:3,type="regressor", data=imoveiscwbav)

        RESET test

data:  price ~ age + parea + tarea + bath + ensuit + garag + plaz +     park + trans + kidca + school + health + bike + barb
 + balc +     elev + fitg + party + categ
RESET = 5.2815, df1 = 38, df2 = 483, p-value < 2.2e-16

>
>
> # F tabelado:
>
> qf(0.95, df1=38, df2=483)
[1] 1.429987
```

Como o F calculado (5.2815) é maior que o F tabelado (1,429987), existe erro de especificação do modelo

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

```
>
> cor(imoveiscwbav[,c("age", "parea", "tarea", "bath", "ensuit", "garag", "plaz", "park",
+                     "trans", "kidca", "school", "health", "bike")],
+     use="complete")
               age       parea       tarea        bath      ensuit        garag         plaz        park       trans
age     1.000000000 -0.23698282 -0.31657436 -0.26037524 -0.41131287 -0.448491700  0.007897195  0.04544181 -0.05481368
parea  -0.236982819  1.00000000  0.83631214  0.68030305  0.59703635  0.603548940 -0.103245172  0.21597209 -0.24465672
tarea  -0.316574355  0.83631214  1.00000000  0.65178059  0.56702436  0.645854936 -0.137222512  0.21515484 -0.20718875
bath   -0.260375235  0.68030305  0.65178059  1.00000000  0.73616242  0.574197743 -0.112992339  0.18807029 -0.16147897
ensuit -0.411312872  0.59703635  0.56702436  0.73616242  1.00000000  0.531151588 -0.083877169  0.04924212 -0.04513610
garag  -0.448491700  0.60354894  0.64585494  0.57419774  0.53115159  1.000000000 -0.080390991  0.15699883 -0.13018970
plaz    0.007897195 -0.10324517 -0.13722251 -0.11299234 -0.08387717 -0.080390991  1.000000000 -0.32462706  0.01029101
park    0.045441811  0.21597209  0.21515484  0.18807029  0.04924212  0.156998830 -0.324627059  1.00000000 -0.29767701
trans  -0.054813684 -0.24465672 -0.20718875 -0.16147897 -0.04513610 -0.130189703  0.010291011 -0.29767701  1.00000000
kidca  -0.007034162  0.23719013  0.26214005  0.23256465  0.15681143  0.131592614 -0.100390573  0.29848179 -0.50283754
school  0.096298011  0.15219562  0.09328829  0.07286173  0.02216520  0.005902974  0.160404566  0.46081487 -0.11388155
health -0.097090851 -0.12890024 -0.13439880 -0.14426728 -0.12274559 -0.022402611  0.009381361 -0.24700544  0.15173518
bike    0.070154638 -0.06615144 -0.06336755 -0.06967850 -0.05244026 -0.068895267  0.222888734 -0.38994669 -0.08404754
              kidca       school       health         bike
age    -0.007034162  0.096298011 -0.097090851  0.07015464
parea   0.237190133  0.152195617 -0.128900236 -0.06615144
tarea   0.262140051  0.093288289 -0.134398800 -0.06336755
bath    0.232564655  0.072861730 -0.144267282 -0.06967850
ensuit  0.156811432  0.022165200 -0.122745589 -0.05244026
garag   0.131592614  0.005902974 -0.022402611 -0.06889527
plaz   -0.100390573  0.160404566  0.009381361  0.22288873
park    0.298481791  0.460814867 -0.247005445 -0.38994669
trans  -0.502837545 -0.113881554  0.151735185 -0.08404754
kidca   1.000000000  0.146596839 -0.349252096  0.18358136
school  0.146596839  1.000000000 -0.373524820  0.07078386
health -0.349252096 -0.373524820  1.000000000 -0.11611012
bike    0.183581360  0.070783858 -0.116110122  1.00000000
> library(car)
> vif(lm(price~age+parea+tarea+bath+ensuit+garag+plaz+park+trans+kidca+school+health+bike,data=imoveiscwbav), type="high-orde
r")
```

Parece que o parea e tarea são correlacionadas pela matriz de correlação. Pelo valor de inflação de variancia nota-se que que o parea e o tarea tem score alto, quase atingindo valor quatro, então estão consideravelmente correlacionados, a sugestão é excluir o tarea. mas serão feitas exclusões após o stepwise.

f) Selecione um modelo pela técnica de stepwise;

```
> resultados <- lm(price~age+parea+tarea+bath+ensuit+garag+plaz+park+trans+kidca+school+health+bike+barb+balc+elev+
+                    fitg+party+categ,
+                  data=imoveiscwbav)
> summary (resultados)

Call:
lm(formula = price ~ age + parea + tarea + bath + ensuit + garag +
    plaz + park + trans + kidca + school + health + bike + barb +
    balc + elev + fitg + party + categ, data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-575422 -132235   -1347   98967 2408789

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1011256.2   141686.1  -7.137 3.21e-12 ***
age             -129.3       25.3  -5.110 4.54e-07 ***
parea          46751.5    14519.8   3.220  0.00136 **
tarea          55178.0     9758.0   5.655 2.58e-08 ***
bath            4079.9    15284.1   0.267  0.78962
ensuit        153095.5    18852.5   8.121 3.37e-15 ***
garag         184403.4    22478.4   8.204 1.84e-15 ***
plaz           53988.2    18080.8   2.986  0.00296 **
park          190747.3    50923.4   3.746  0.00020 ***
trans         -89961.3    44710.8  -2.012  0.04473 *
kidca         -27557.6    37516.0  -0.735  0.46294
school        -67472.8    62015.7  -1.088  0.27710
health         16456.8    57633.3   0.286  0.77534
bike          -45288.4    61323.6  -0.739  0.46053
barb          -36985.5    23387.1  -1.581  0.11438
balc           79093.6    25959.3   3.047  0.00243 **
elev         -106901.9    26176.1  -4.084 5.12e-05 ***
fitg          139061.0    29377.3   4.734 2.85e-06 ***
party          23647.6    29530.0   0.801  0.42361
categ         259957.1    56205.5   4.625 4.73e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 237200 on 521 degrees of freedom
Multiple R-squared:  0.7967,    Adjusted R-squared:  0.7893
F-statistic: 107.5 on 19 and 521 DF,  p-value: < 2.2e-16
```

```
> step <- stepwise(resultados, direction= 'backward/forward', criterion ='AIC')

Direction:  backward/forward
Criterion:  AIC

Start:  AIC=13411.33
price ~ age + parea + tarea + bath + ensuit + garag + plaz +
    park + trans + kidca + school + health + bike + barb + balc +
    elev + fitg + party + categ

          Df  Sum of Sq        RSS   AIC
- bath     1 4.0103e+09 2.9326e+13 13409
- health   1 4.5889e+09 2.9327e+13 13409
- kidca    1 3.0367e+10 2.9353e+13 13410
- bike     1 3.0696e+10 2.9353e+13 13410
- party    1 3.6092e+10 2.9358e+13 13410
- school   1 6.6621e+10 2.9389e+13 13411
<none>                  2.9322e+13 13411
- barb     1 1.4076e+11 2.9463e+13 13412
- trans    1 2.2785e+11 2.9550e+13 13414
- plaz     1 5.0179e+11 2.9824e+13 13418
- balc     1 5.2246e+11 2.9845e+13 13419
- parea    1 5.8349e+11 2.9906e+13 13420
- park     1 7.8966e+11 3.0112e+13 13424
- elev     1 9.3869e+11 3.0261e+13 13426
- categ    1 1.2039e+12 3.0526e+13 13431
- fitg     1 1.2611e+12 3.0583e+13 13432
- age      1 1.4693e+12 3.0792e+13 13436
- tarea    1 1.7996e+12 3.1122e+13 13442
- ensuit   1 3.7115e+12 3.3034e+13 13474
- garag    1 3.7876e+12 3.3110e+13 13475

Step:  AIC=13409.4
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + kidca + school + health + bike + barb + balc + elev +
    fitg + party + categ

          Df  Sum of Sq        RSS   AIC
- health   1 4.4490e+09 2.9331e+13 13408
- kidca    1 2.9571e+10 2.9356e+13 13408
- bike     1 3.1098e+10 2.9357e+13 13408
- party    1 3.8460e+10 2.9365e+13 13408
- school   1 6.7651e+10 2.9394e+13 13409
<none>                  2.9326e+13 13409
- barb     1 1.3893e+11 2.9465e+13 13410
+ bath     1 4.0103e+09 2.9322e+13 13411
- trans    1 2.2931e+11 2.9556e+13 13412
- plaz     1 5.0337e+11 2.9830e+13 13417
- balc     1 5.2621e+11 2.9852e+13 13417
- parea    1 6.1896e+11 2.9945e+13 13419
- park     1 8.0168e+11 3.0128e+13 13422
- elev     1 9.3556e+11 3.0262e+13 13424
- categ    1 1.2047e+12 3.0531e+13 13429
- fitg     1 1.2589e+12 3.0585e+13 13430
- age      1 1.4838e+12 3.0810e+13 13434
- tarea    1 1.8270e+12 3.1153e+13 13440
- garag    1 3.8980e+12 3.3224e+13 13475
- ensuit   1 5.3272e+12 3.4653e+13 13498
```

```
Step:  AIC=13407.48
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + kidca + school + bike + barb + balc + elev + fitg +
    party + categ

          Df  Sum of Sq        RSS   AIC
- bike     1 3.2891e+10 2.9364e+13 13406
- kidca    1 3.7660e+10 2.9368e+13 13406
- party    1 4.0511e+10 2.9371e+13 13406
- school   1 8.2675e+10 2.9413e+13 13407
<none>                  2.9331e+13 13408
- barb     1 1.3826e+11 2.9469e+13 13408
+ health   1 4.4490e+09 2.9326e+13 13409
+ bath     1 3.8704e+09 2.9327e+13 13409
- trans    1 2.3217e+11 2.9563e+13 13410
- plaz     1 5.0507e+11 2.9836e+13 13415
- balc     1 5.2189e+11 2.9853e+13 13415
- parea    1 6.2309e+11 2.9954e+13 13417
- park     1 7.9895e+11 3.0130e+13 13420
- elev     1 9.3610e+11 3.0267e+13 13422
- categ    1 1.2004e+12 3.0531e+13 13427
- fitg     1 1.2575e+12 3.0588e+13 13428
- age      1 1.5159e+12 3.0847e+13 13433
- tarea    1 1.8227e+12 3.1153e+13 13438
- garag    1 3.9163e+12 3.3247e+13 13473
- ensuit   1 5.3684e+12 3.4699e+13 13496

Step:  AIC=13406.09
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + kidca + school + barb + balc + elev + fitg + party +
    categ

          Df  Sum of Sq        RSS   AIC
- party    1 3.6727e+10 2.9400e+13 13405
- kidca    1 6.6838e+10 2.9430e+13 13405
<none>                  2.9364e+13 13406
- school   1 1.3381e+11 2.9497e+13 13406
- barb     1 1.6076e+11 2.9524e+13 13407
+ bike     1 3.2891e+10 2.9331e+13 13408
+ health   1 6.2424e+09 2.9357e+13 13408
+ bath     1 4.2513e+09 2.9359e+13 13408
- trans    1 2.1733e+11 2.9581e+13 13408
- plaz     1 5.0627e+11 2.9870e+13 13413
- balc     1 5.1341e+11 2.9877e+13 13414
- parea    1 6.4815e+11 3.0012e+13 13416
- elev     1 9.4566e+11 3.0309e+13 13421
- categ    1 1.2264e+12 3.0590e+13 13426
- park     1 1.2947e+12 3.0658e+13 13427
- fitg     1 1.3032e+12 3.0667e+13 13428
- age      1 1.5804e+12 3.0944e+13 13432
- tarea    1 1.8139e+12 3.1177e+13 13436
- garag    1 3.8921e+12 3.3256e+13 13471
- ensuit   1 5.3953e+12 3.4759e+13 13495
```

```
Step:  AIC=13404.77
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + kidca + school + barb + balc + elev + fitg + categ

          Df  Sum of Sq        RSS   AIC
- kidca    1 6.2555e+10 2.9463e+13 13404
<none>                  2.9400e+13 13405
- school   1 1.3636e+11 2.9537e+13 13405
- barb     1 1.5616e+11 2.9556e+13 13406
+ party    1 3.6727e+10 2.9364e+13 13406
+ bike     1 2.9108e+10 2.9371e+13 13406
+ health   1 8.3752e+09 2.9392e+13 13407
+ bath     1 6.5661e+09 2.9394e+13 13407
- trans    1 2.1281e+11 2.9613e+13 13407
- plaz     1 5.3543e+11 2.9936e+13 13412
- balc     1 6.2350e+11 3.0024e+13 13414
- parea    1 6.5043e+11 3.0051e+13 13415
- elev     1 9.3541e+11 3.0336e+13 13420
- categ    1 1.2883e+12 3.0689e+13 13426
- park     1 1.2959e+12 3.0696e+13 13426
- age      1 1.5926e+12 3.0993e+13 13431
- fitg     1 1.7557e+12 3.1156e+13 13434
- tarea    1 1.8491e+12 3.1249e+13 13436
- garag    1 3.8819e+12 3.3282e+13 13470
- ensuit   1 5.3846e+12 3.4785e+13 13494

Step:  AIC=13403.92
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + school + barb + balc + elev + fitg + categ

          Df  Sum of Sq        RSS   AIC
<none>                  2.9463e+13 13404
- school   1 1.4668e+11 2.9610e+13 13405
- trans    1 1.5192e+11 2.9615e+13 13405
+ kidca    1 6.2555e+10 2.9400e+13 13405
+ bike     1 5.6072e+10 2.9407e+13 13405
- barb     1 1.6682e+11 2.9630e+13 13405
+ party    1 3.2445e+10 2.9430e+13 13405
+ health   1 2.4191e+10 2.9439e+13 13406
+ bath     1 4.8823e+09 2.9458e+13 13406
- plaz     1 5.5372e+11 3.0017e+13 13412
- balc     1 6.3695e+11 3.0100e+13 13414
- parea    1 6.5574e+11 3.0119e+13 13414
- elev     1 9.2837e+11 3.0391e+13 13419
- categ    1 1.2392e+12 3.0702e+13 13424
- park     1 1.2586e+12 3.0721e+13 13424
- age      1 1.6042e+12 3.1067e+13 13431
- fitg     1 1.7060e+12 3.1169e+13 13432
- tarea    1 1.8031e+12 3.1266e+13 13434
- garag    1 3.9794e+12 3.3442e+13 13470
- ensuit   1 5.3259e+12 3.4789e+13 13492
> step

Call:
lm(formula = price ~ age + parea + tarea + ensuit + garag + plaz +
    park + trans + school + barb + balc + elev + fitg + categ,
    data = imoveiscwbav)
```

```
Call:
lm(formula = price ~ age + parea + tarea + ensuit + garag + plaz +
    park + trans + school + barb + balc + elev + fitg + categ,
    data = imoveiscwbav)

Coefficients:
(Intercept)          age        parea        tarea       ensuit        garag
 -1064753.1       -131.8      48645.0      54666.4     154017.8     186255.6
       plaz         park        trans       school         barb         balc
   56407.1     207115.1     -64673.1     -91266.3     -39758.7      84156.9
       elev         fitg        categ
  -99341.2     147195.1     259282.7
```

Melhor Modelo

```
> resultados <- lm(price ~ age + parea + tarea + ensuit + garag + plaz +
+                    park + trans + bike + barb + balc + elev + fitg + categ, data = imoveiscwbav)
> summary(resultados)

Call:
lm(formula = price ~ age + parea + tarea + ensuit + garag + plaz +
    park + trans + bike + barb + balc + elev + fitg + categ,
    data = imoveiscwbav)

Residuals:
    Min      1Q  Median      3Q     Max
-553655 -145325   -1028  106340 2394627

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.046e+06  1.252e+05  -8.352 5.98e-16 ***
age         -1.302e+02  2.476e+01  -5.258 2.12e-07 ***
parea        4.591e+04  1.420e+04   3.234 0.001297 **
tarea        5.525e+04  9.650e+03   5.725 1.74e-08 ***
ensuit       1.543e+05  1.580e+04   9.763  < 2e-16 ***
garag        1.894e+05  2.200e+04   8.610  < 2e-16 ***
plaz         4.858e+04  1.681e+04   2.889 0.004024 **
park         1.480e+05  3.955e+04   3.743 0.000202 ***
trans       -8.332e+04  3.999e+04  -2.084 0.037684 *
bike        -8.097e+04  5.513e+04  -1.469 0.142511
barb        -3.314e+04  2.316e+04  -1.431 0.153088
balc         8.671e+04  2.495e+04   3.475 0.000553 ***
elev        -9.766e+04  2.443e+04  -3.998 7.30e-05 ***
fitg         1.414e+05  2.640e+04   5.354 1.29e-07 ***
categ        2.699e+05  5.435e+04   4.966 9.26e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 236800 on 526 degrees of freedom
Multiple R-squared:  0.7955,    Adjusted R-squared:  0.7901
F-statistic: 146.2 on 14 and 526 DF,  p-value: < 2.2e-16
```

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

H0 - homocedástico(variáveis constantes)
HA - Heterocedástico(variáveis não constantes)

Breusch-Pagan

```
> bptest(price ~ age + parea + tarea + ensuit + garag + plaz +
+            park + trans + bike + barb + balc + elev + fitg + categ,studentize=FALSE, data=imoveiscwbav)

        Breusch-Pagan test

data:  price ~ age + parea + tarea + ensuit + garag + plaz + park +     trans + bike + barb + balc + elev + fitg + categ
BP = 282.15, df = 14, p-value < 2.2e-16
```

Qui-quadratico

```
> qchisq(0.95, df=14, lower.tail = TRUE)
[1] 23.68479
```

O valor de 290 está muito acima do tabelado, então rejeita-se a hipótese de hocedascidade. Será feito regressão robusta.

```
> resultrob <- lmRob(price ~ age + parea + tarea + ensuit + garag + plaz +
+                      park + trans + bike + barb + balc + elev + fitg + categ,data=imoveiscwbav)
```

h) Obtenha os indicadores de desempenho do modelo;

```
> AIC(resultados)
[1] 14941.68
> BIC(resultados)
[1] 15010.38
> library(AICcmodavg)
> AICc(resultados)
[1] 14942.72
> library(performance)
> model_performance(resultados)
# Indices of model performance

AIC       |       BIC |    R2 | R2 (adj.) |      RMSE |     Sigma
-----------------------------------------------------------------
14912.255 | 15002.416 | 0.810 |     0.803 | 2.251e+05 | 2.294e+05
> model_performance(resultrob)
# Indices of model performance

R2    |      RMSE |     Sigma
----------------------------
0.791 | 2.363e+05 | 2.396e+05
```

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

```
> confint(resultados, level = 0.95)
                    2.5 %        97.5 %
(Intercept) -1291817.4425 -799842.52225
age              -178.8353     -81.55662
parea           18022.3061   73795.65298
tarea           36288.4021   74202.61980
ensuit         123220.9512  185299.01289
garag          146177.6906  232606.90880
plaz            15544.6432   81607.73100
park            70328.3483  225731.71107
trans         -161872.4267   -4760.74702
bike          -189265.0873   27331.53781
barb           -78648.8733   12362.97252
balc            37691.7022  135723.22050
elev          -145646.3242  -49675.84270
fitg            89483.5383  193218.38150
categ          163122.9125  376665.68768
```

```
> confint(resultrob, level=0.95)
                   2.5 %      97.5 %
(Intercept) -647091.1924 -320777.322
age            -8746.0011   -5840.952
parea           1733.4274    3628.755
tarea            991.0716    2107.736
ensuit        108310.7009  160089.136
garag         113680.5160  179170.787
plaz          188620.6910  454926.198
park          -67028.1510    1928.786
trans           6647.7359   65755.211
bike         -176836.4768  -18368.795
barb          -42461.7478   26605.673
balc           17244.2960   91460.756
elev         -102680.7428  -27299.923
fitg           60802.5216  140534.528
categ         289785.2257  452885.540
```

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

```
> ################ Fazendo predições no modelos OLS
> #Para:
> #age = 5 anos = 1.609438
> log(5)
[1] 1.609438
> #educ = 150 parea = 5.010635
> log(150)
[1] 5.010635
> #earns = 190 tarea = 5.247024
> log(250)
[1] 5.521461
> # plaz
> log(0.08)
[1] -2.525729
> #kidlt6 = 1 --> tem filhos com menos de 6 anos
> predict(object = resultrob,
+         data.frame(age=1.60, parea=5.01, tarea=5.24, ensuit = 1, garag = 1, plaz = -2.525729,
+                    park = -2.525729, trans = -2.525729, bike = -2.525729, categ = 1,
+                    barb=0, balc=0, elev=0, fitg=0))
       1
-397509.9
>
> #o salario por hora é:
> exp(14.11198)
[1] 1345101
>
> anova.lmRob(resultrob)
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left

Terms added sequentially (first to last)

            Chisq Df RobustF     Pr(F)
(Intercept)     1
age             1  205.08 < 2.2e-16 ***
parea           1  410.03 < 2.2e-16 ***
tarea           1   84.91 < 2.2e-16 ***
ensuit          1   77.37 < 2.2e-16 ***
garag           1   58.07 8.105e-15 ***
plaz            1   11.58 0.0005245 ***
park            1   14.02 0.0001358 ***
trans           1    1.72 0.1813986
bike            1   12.32 0.0003474 ***
barb            1    0.07 0.7886646
balc            1    5.49 0.0169267 *
elev            1    5.11 0.0212121 *
fitg            1   28.85 4.414e-08 ***
categ           1   55.34 3.431e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Warning messages:
1: In lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control,  :
  Max iteration for refinement reached.
2: In lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control,  :
  Max iteration for refinement reached.
3: In lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control,  :
  Max iteration for refinement reached.
4: In lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control,  :
  Max iteration for refinement reached.
```

```
> anova(resultados)
Analysis of Variance Table

Response: price
           Df     Sum Sq    Mean Sq  F value    Pr(>F)
age         1 4.1181e+13 4.1181e+13 782.5857 < 2.2e-16 ***
parea       1 4.9625e+13 4.9625e+13 943.0548 < 2.2e-16 ***
tarea       1 7.5864e+12 7.5864e+12 144.1677 < 2.2e-16 ***
bath        1 4.3846e+12 4.3846e+12  83.3224 < 2.2e-16 ***
ensuit      1 2.8791e+12 2.8791e+12  54.7126 5.616e-13 ***
garag       1 3.9279e+12 3.9279e+12  74.6439 < 2.2e-16 ***
plaz        1 7.8541e+10 7.8541e+10   1.4925  0.222375
park        1 1.9113e+12 1.9113e+12  36.3214 3.168e-09 ***
trans       1 2.2340e+10 2.2340e+10   0.4245  0.514964
kidca       1 7.6335e+10 7.6335e+10   1.4506  0.228973
school      1 5.2679e+10 5.2679e+10   1.0011  0.317514
health      1 2.5533e+05 2.5533e+05   0.0000  0.998243
bike        1 2.9858e+11 2.9858e+11   5.6741  0.017575 *
barb        1 5.9962e+10 5.9962e+10   1.1395  0.286254
balc        1 8.5169e+11 8.5169e+11  16.1850 6.596e-05 ***
elev        1 4.1658e+11 4.1658e+11   7.9165  0.005084 **
fitg        1 1.9156e+12 1.9156e+12  36.4025 3.047e-09 ***
party       1 1.7862e+11 1.7862e+11   3.3944  0.065986 .
categ       1 1.3613e+12 1.3613e+12  25.8693 5.106e-07 ***
Residuals 521 2.7416e+13 5.2622e+10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```