Estatística Aplicada I

Segunda Lista de Exercícios

Estudante: Clístenes Grizafis Bento

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

a) Estimar um modelo preliminar e apresentar os resultados;

In [197…
```
#Carreando bibliotecas
load("imoveiscwbav.RData")
gc()
```

	used	(Mb)	gc trigger	(Mb)	max used	(Mb)
Ncells	3712662	198.3	7625532	407.3	7625532	407.3
Vcells	6892518	52.6	12267951	93.6	10759652	82.1

In [198…
```
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
```

In [199…

```
imoveiscwbav
```

price	age	parea	tarea	bath	ensuit	garag	plaz	park	trans	kidca	school	health	bike	barb	balc	elev	fitg
1100000	15	150	190	4	1	2	0.08058169	0.7132806	2.3862709	1.4109813	0.9028108	0.4146473	0.21319266	0	1	0	1
895000	11	165	210	4	1	2	0.16635098	0.6983694	2.2463043	1.8625914	0.9355790	0.2569533	0.23255291	0	0	0	0
2513600	2	146	275	4	3	3	0.05607530	1.3129824	2.6314112	1.5914926	0.4517910	0.2321598	0.29709268	0	1	0	1
755000	25	163	238	3	1	2	0.32159391	2.1099578	2.1387003	1.6215857	0.4478709	0.6848450	0.34714701	0	0	0	0
1099000	1	107	189	3	1	2	0.14663511	1.0175299	1.7978931	1.2572430	0.8841994	0.2990089	0.77876451	0	0	1	1
475000	31	96	124	2	1	1	0.12615199	1.9700244	0.9947105	1.0973424	0.3915381	0.2799555	0.29670284	0	0	0	0
463900	2	75	90	2	1	1	0.22649932	2.6027009	1.4633611	1.0064518	0.2742137	0.6350527	0.11696390	0	0	0	0
1320000	1	122	227	3	1	2	0.19124408	1.1613287	2.8183332	1.3903009	0.3598062	0.3802676	0.14616086	1	1	1	0
550000	11	63	87	2	1	1	0.46403846	1.9965809	1.8184778	1.1122639	0.4473747	0.6192844	0.73284788	0	0	0	0
831000	1	97	180	2	1	2	0.19130703	2.1665835	2.3593991	0.6772416	0.6670535	0.7362139	0.19716810	1	1	1	1
979000	3	92	130	2	1	3	0.35417661	2.0669094	2.2013961	1.5627364	0.5104637	0.7312597	0.30117480	0	0	0	0
1817417	3	138	253	3	3	2	0.39193020	2.1114371	1.7348651	1.6228450	0.5825087	0.0959644	0.59292218	0	0	0	0
2892000	3	199	405	5	3	3	0.11616112	0.7025960	2.4255204	1.5600436	0.7921303	0.2551667	0.09454164	0	1	0	0
1050000	11	150	255	3	1	2	0.08058169	0.7132806	2.3862709	1.4109813	0.9028108	0.4146473	0.21319266	0	1	0	1
959000	3	92	130	2	1	2	0.35417661	2.0669094	2.2013961	1.5627364	0.5104637	0.7312597	0.30117480	0	0	0	0
490000	20	86	148	3	1	1	0.12022662	2.1309628	1.8639108	1.5859712	0.3758302	0.3248843	0.51606304	0	0	0	0
1548000	3	139	241	5	3	2	0.35713365	2.3320754	1.4810736	1.4557371	0.6223256	0.1757756	0.53096446	1	1	0	1
2835000	3	199	284	5	3	3	0.11587989	0.7088692	2.4374977	1.5430888	0.7865406	0.2662776	0.07732327	0	1	0	1
850000	38	184	281	2	1	2	0.30068165	2.3396719	1.4572900	1.4779062	0.5602813	0.1979433	0.58670117	1	1	0	0
995000	16	164	206	4	3	2	0.07345642	1.4411875	2.5190436	1.6050825	0.4520683	0.3310010	0.42252675	1	1	1	0
2017000	3	150	215	5	3	3	0.11587989	0.7088692	2.4374977	1.5430888	0.7865406	0.2662776	0.07732327	0	1	0	1
1692000	3	139	241	4	3	3	0.35713365	2.3320754	1.4810736	1.4557371	0.6223256	0.1757756	0.53096446	1	1	0	1
1720000	3	150	215	5	3	3	0.11587989	0.7088692	2.4374977	1.5430888	0.7865406	0.2662776	0.07732327	0	1	0	1
530000	22	78	97	2	1	1	0.21268581	0.8530269	2.3086127	1.2707697	0.8224284	0.3612718	0.18452322	0	0	0	0

price	age	parea	tarea	bath	ensuit	garag	plaz	park	trans	kidca	school	health	bike	barb	balc	elev	fitg
1850000	1	158	297	5	2	3	0.34226296	2.0804271	2.2631276	1.4902507	0.5584116	0.8132937	0.21831460	0	0	0	0
1079000	1	107	196	3	1	2	0.14663511	1.0175299	1.7978931	1.2572430	0.8841994	0.2990089	0.77876451	1	1	1	0
1485603	3	155	241	3	3	2	0.39193020	2.1114371	1.7348651	1.6228450	0.5825087	0.0959644	0.59292218	0	0	0	0
979000	16	145	215	4	1	2	0.08318506	0.8471284	2.4272793	1.7265898	0.7574189	0.3160069	0.19945269	1	0	0	1
1358000	3	121	185	3	1	2	0.17224844	2.2469464	2.1452308	1.5680154	0.4294937	0.7943064	0.29570670	0	1	0	1
730000	13	110	128	2	1	2	0.23733018	1.9439387	0.7801857	0.6116902	0.4370783	0.7072616	0.25919896	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1864000	3	155	254	4	3	2	0.3559855	2.331109	1.481412	1.4574720	0.6208978	0.1765076	0.53328193	0	1	0	1
1494000	3	139	241	4	3	2	0.3559855	2.331109	1.481412	1.4574720	0.6208978	0.1765076	0.53328193	0	1	0	1
1850000	3	158	258	5	3	3	0.3370544	2.084975	2.256744	1.4962215	0.5518671	0.8087530	0.22256929	1	1	1	1
1320000	1	122	228	3	1	2	0.1974011	1.160932	2.814845	1.3757335	0.3537807	0.3958252	0.15562187	1	0	0	0
1920000	3	132	228	4	2	3	0.1449806	1.419008	2.688781	1.1069239	0.1068977	0.6009452	0.72807238	1	0	0	1
1180000	1	122	225	3	1	2	0.1936256	1.160322	2.818933	1.3866845	0.3589145	0.3843136	0.14791330	0	1	0	0
340000	34	79	115	2	1	1	0.1361207	2.531448	1.944107	1.3111943	0.3004588	0.8290731	0.55060481	0	0	0	0
750000	12	162	213	3	1	3	0.1519592	2.216822	1.376249	1.2340717	0.3257275	0.1994018	0.81871366	1	1	0	0
520000	10	135	190	2	1	1	0.4983964	1.874185	1.952006	1.6964949	0.7505681	0.3313860	0.81353794	0	0	0	0
595000	17	131	180	3	1	2	0.2233063	2.396326	1.234467	0.2456014	0.5645545	1.1974253	1.02706436	1	0	0	0
740000	12	162	259	3	1	3	0.3173683	2.235832	1.541227	1.2977872	0.4955706	0.2956851	0.81095804	1	0	0	0
949000	1	147	217	4	3	2	0.2328303	2.101013	2.413553	1.3266400	0.6257842	0.6826416	0.03851381	1	0	0	0
820000	10	140	155	3	1	2	0.4354084	1.984615	2.299085	1.4777980	0.5813712	0.7981299	0.26320560	1	1	0	0
858980	1	148	200	2	3	2	0.0835254	2.190340	1.925189	1.5566143	0.2917574	0.4782124	0.51079306	1	1	0	0
950000	1	179	188	5	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	1	0	0	0
950013	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950000	1	179	188	5	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	1	0	0	0

price	age	parea	tarea	bath	ensuit	garag	plaz	park	trans	kidca	school	health	bike	barb	balc	elev	fitg
950000	1	179	188	3	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950040	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950001	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	1	0	0
950012	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950008	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950003	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
950000	1	179	188	4	3	2	0.2485592	2.453074	2.529130	1.2673058	0.8699518	0.2808784	0.38352472	0	0	0	0
560000	41	129	140	3	0	2	0.4060568	1.966891	2.150560	0.3197310	0.8484569	0.4715769	0.70321453	0	0	0	0
1814400	1	152	300	5	3	3	0.1187995	1.302387	2.818119	1.2964524	0.2122319	0.4733070	0.29201095	1	1	1	1
875000	10	125	126	3	1	2	0.3377776	2.549630	1.391650	0.8775522	0.3943746	0.7519699	0.22808992	1	0	0	0
1003300	1	122	237	3	1	2	0.1925611	1.160392	2.817936	1.3894167	0.3600906	0.3814008	0.14618764	0	1	1	0
900000	15	125	210	4	1	3	0.1102672	1.369616	2.408082	1.6823833	0.9654204	0.3861878	0.97059600	1	1	0	0
595000	17	131	180	2	1	2	0.2233063	2.396326	1.234467	0.2456014	0.5645545	1.1974253	1.02706436	1	1	0	0

b) Testar as variáveis para formulação do modelo;

In [200…

```
library("PanJen")

#teste com 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))

PanJenage<-fform(imoveiscwbav,"age",formBase)
```

```
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
               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"
```

smoothing vencedora, incluir age

In [201…

```
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))

PanJenparea<-fform(imoveiscwbav,"parea",formBase)
```

```
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
               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"
```

sqr(x) vencedora, incluir parea

In [202…

```
library(timetk)

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))

PanJentarea<-fform(imoveiscwbav,"tarea",formBase)
```

```
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
               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"
```

smoothing vencedora, incluir tarea

In [203…

```
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))

PanJenplaz<-fform(imoveiscwbav,"plaz",formBase)
```

```
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
               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"
```

base vencedora, excluir plaz

In [204…

```
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))

PanJenpark<-fform(imoveiscwbav,"park",formBase)
```

```
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
               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"
```

smoothing vencedora, incluir park

In [205…

```
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))

PanJentrans<-fform(imoveiscwbav,"trans",formBase)
```

```
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
               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"
```

base vencedora, excluir trans

In [206…

```
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))

PanJenkidca<-fform(imoveiscwbav,"kidca",formBase)
```

```
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
               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"
```

base vencedora, excluir kidca

In [207…

```
ormBase<-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))

PanJenschool<-fform(imoveiscwbav,"school",formBase)
```

```
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
```

```
               AIC      BIC ranking (BIC)
base      14910.35 14996.22           1.5
x         14910.35 14996.22           1.5
1/x       14908.25 14998.41           3.0
log(x)    14911.83 15001.99           4.0
x^2       14912.08 15002.24           5.0
sqr(x)    14912.25 15002.42           6.5
x+x^2     14912.25 15002.42           6.5
smoothing 14900.87 15017.99           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"
```

base vencedora, excluir school

In [208…

```
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))

PanJenhealth<-fform(imoveiscwbav,"health",formBase)
```

```
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
               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"
```

base vencedora, excluir health

In [209…

```
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))

PanJenbike<-fform(imoveiscwbav,"bike",formBase)
```

```
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
               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"
```

base vencedora, excluir bike

In [210…

```
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
```

c) Verifique a presença de outliers;

In [211…

```
library (carData)
library(car)
```

```
outlierTest(resultados)
```

```
     rstudent unadjusted p-value Bonferroni p
393 12.131725         5.3325e-30   2.8849e-27
13   4.302072         2.0218e-05   1.0938e-02
```

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

In [212…

```
library (zoo)
library (lmtest)

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 + b
alc +     elev + fitg + party + categ
RESET = 4.3391, df1 = 38, df2 = 483, p-value = 9.761e-15
```

F tabelado

In [213…

```
qf(0.95, df1=38, df2=483)
```

1.4299873657331

Como o F calculado (4.3391) é maior que o F tabelado (1,4299873657331), existe erro de especificação do modelo

Deixaremos para remover as variáveis posteriormente

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

In [214…

```
cor(imoveiscwbav[,c("age", "parea", "tarea", "bath", "ensuit", "garag", "plaz", "park",
                    "trans", "kidca", "school", "health", "bike")],
    use="complete")

library(car)
```

```
vif(lm(price~age+parea+tarea+bath+ensuit+garag+plaz+park+trans+kidca+school+health+bike,data=imoveiscwbav), type="high-order")
```

	age	parea	tarea	bath	ensuit	garag	plaz	park	trans	kidca	school
age	1.00000000	-0.23046283	-0.33295431	-0.26832783	-0.44532055	-0.473613016	0.02522293	-0.03471480	-0.01343249	0.01027489	0.084734049
parea	-0.23046283	1.00000000	0.81435888	0.67748403	0.59932048	0.592953262	-0.06585725	-0.21091786	0.28699371	0.23650295	0.161190207
tarea	-0.33295431	0.81435888	1.00000000	0.64462760	0.56271720	0.639279714	-0.09211596	-0.23195234	0.23256402	0.25872055	0.094233700
bath	-0.26832783	0.67748403	0.64462760	1.00000000	0.73616242	0.574197743	-0.06282817	-0.16996914	0.16864749	0.23256465	0.072861730
ensuit	-0.44532055	0.59932048	0.56271720	0.73616242	1.00000000	0.531151588	-0.04448629	-0.02382108	0.06323091	0.15681143	0.022165200
garag	-0.47361302	0.59295326	0.63927971	0.57419774	0.53115159	1.000000000	-0.07992508	-0.16205861	0.17069446	0.13159261	0.005902974
plaz	0.02522293	-0.06585725	-0.09211596	-0.06282817	-0.04448629	-0.079925081	1.00000000	0.31662772	-0.15779668	-0.08194894	0.184979230
park	-0.03471480	-0.21091786	-0.23195234	-0.16996914	-0.02382108	-0.162058611	0.31662772	1.00000000	-0.50306548	-0.27488761	-0.378827033
trans	-0.01343249	0.28699371	0.23256402	0.16864749	0.06323091	0.170694456	-0.15779668	-0.50306548	1.00000000	0.44121318	0.082776136
kidca	0.01027489	0.23650295	0.25872055	0.23256465	0.15681143	0.131592614	-0.08194894	-0.27488761	0.44121318	1.00000000	0.146596839
school	0.08473405	0.16119021	0.09423370	0.07286173	0.02216520	0.005902974	0.18497923	-0.37882703	0.08277614	0.14659684	1.000000000
health	-0.08768266	-0.13449226	-0.13228191	-0.14426728	-0.12274559	-0.022402611	-0.01211491	0.22422407	-0.06120773	-0.34925210	-0.373524820
bike	0.07658592	-0.07005757	-0.07144375	-0.06967850	-0.05244026	-0.068895267	0.27145206	0.32559215	-0.02767855	0.18358136	0.070783858

age	1.54851994963986
parea	3.86401672700158
tarea	3.60635331708673
bath	2.9552961488982
ensuit	2.77374551093955
garag	2.15910416518495
plaz	1.2934320995796
park	2.32196037369952
trans	1.75587021856237
kidca	1.55435185796771
school	1.59261524590408
health	1.36503443451438
bike	1.35994265365145

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;

In [215…

```
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)

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

```
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
```

```
Direction:  backward/forward
Criterion:  AIC

Start:  AIC=13374.96
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
- health  1 2.4678e+07 2.7416e+13 13373
- school  1 1.0436e+09 2.7417e+13 13373
- kidca   1 4.7205e+09 2.7421e+13 13373
- bath    1 4.3284e+10 2.7459e+13 13374
- trans   1 7.2369e+10 2.7488e+13 13374
- party   1 8.6251e+10 2.7502e+13 13375
<none>                 2.7416e+13 13375
- bike    1 1.2339e+11 2.7539e+13 13375
- barb    1 1.9875e+11 2.7615e+13 13377
- park    1 2.8722e+11 2.7703e+13 13379
- plaz    1 2.9847e+11 2.7715e+13 13379
- balc    1 3.5048e+11 2.7767e+13 13380
- parea   1 9.0807e+11 2.8324e+13 13391
- fitg    1 9.8402e+11 2.8400e+13 13392
- elev    1 1.0269e+12 2.8443e+13 13393
- categ   1 1.3613e+12 2.8777e+13 13399
- tarea   1 1.8429e+12 2.9259e+13 13408
- ensuit  1 2.4231e+12 2.9839e+13 13419
- age     1 3.0758e+12 3.0492e+13 13430
- garag   1 3.2011e+12 3.0617e+13 13433

Step:  AIC=13372.96
price ~ age + parea + tarea + bath + ensuit + garag + plaz +
    park + trans + kidca + school + bike + barb + balc + elev +
    fitg + party + categ

         Df  Sum of Sq        RSS   AIC
- school  1 1.2045e+09 2.7417e+13 13371
- kidca   1 4.9256e+09 2.7421e+13 13371
- bath    1 4.3265e+10 2.7459e+13 13372
- trans   1 7.4224e+10 2.7490e+13 13372
- party   1 8.7005e+10 2.7503e+13 13373
<none>                 2.7416e+13 13373
- bike    1 1.2468e+11 2.7541e+13 13373
- barb    1 1.9880e+11 2.7615e+13 13375
```

```
+ health  1 2.4678e+07 2.7416e+13 13375
- park    1 2.8901e+11 2.7705e+13 13377
- plaz    1 2.9875e+11 2.7715e+13 13377
- balc    1 3.5200e+11 2.7768e+13 13378
- parea   1 9.0829e+11 2.8324e+13 13389
- fitg    1 9.8400e+11 2.8400e+13 13390
- elev    1 1.0271e+12 2.8443e+13 13391
- categ   1 1.3636e+12 2.8780e+13 13397
- tarea   1 1.8438e+12 2.9260e+13 13406
- ensuit  1 2.4419e+12 2.9858e+13 13417
- age     1 3.0967e+12 3.0513e+13 13429
- garag   1 3.2076e+12 3.0624e+13 13431

Step:  AIC=13370.99
price ~ age + parea + tarea + bath + ensuit + garag + plaz +
    park + trans + kidca + bike + barb + balc + elev + fitg +
    party + categ

         Df  Sum of Sq        RSS   AIC
- kidca   1 4.6214e+09 2.7422e+13 13369
- bath    1 4.3625e+10 2.7461e+13 13370
- trans   1 8.4218e+10 2.7501e+13 13371
- party   1 8.7269e+10 2.7505e+13 13371
<none>                2.7417e+13 13371
- bike    1 1.3542e+11 2.7553e+13 13372
- barb    1 1.9882e+11 2.7616e+13 13373
+ school  1 1.2045e+09 2.7416e+13 13373
+ health  1 1.8560e+08 2.7417e+13 13373
- plaz    1 3.1728e+11 2.7735e+13 13375
- balc    1 3.5541e+11 2.7773e+13 13376
- park    1 3.7232e+11 2.7790e+13 13376
- parea   1 9.1638e+11 2.8334e+13 13387
- fitg    1 9.9202e+11 2.8409e+13 13388
- elev    1 1.0264e+12 2.8444e+13 13389
- categ   1 1.4389e+12 2.8856e+13 13397
- tarea   1 1.8517e+12 2.9269e+13 13404
- ensuit  1 2.4449e+12 2.9862e+13 13415
- age     1 3.0980e+12 3.0515e+13 13427
- garag   1 3.2292e+12 3.0647e+13 13429

Step:  AIC=13369.08
price ~ age + parea + tarea + bath + ensuit + garag + plaz +
    park + trans + bike + barb + balc + elev + fitg + party +
    categ
```

```
         Df  Sum of Sq        RSS   AIC
- bath    1 4.5750e+10 2.7468e+13 13368
- party   1 8.8389e+10 2.7510e+13 13369
<none>                2.7422e+13 13369
- trans   1 1.1228e+11 2.7534e+13 13369
- bike    1 1.3202e+11 2.7554e+13 13370
+ kidca   1 4.6214e+09 2.7417e+13 13371
- barb    1 1.9928e+11 2.7621e+13 13371
+ school  1 9.0034e+08 2.7421e+13 13371
+ health  1 4.5655e+07 2.7422e+13 13371
- plaz    1 3.1415e+11 2.7736e+13 13373
- balc    1 3.5223e+11 2.7774e+13 13374
- park    1 3.8300e+11 2.7805e+13 13375
- parea   1 9.1582e+11 2.8338e+13 13385
- fitg    1 1.0166e+12 2.8439e+13 13387
- elev    1 1.0304e+12 2.8452e+13 13387
- categ   1 1.4854e+12 2.8907e+13 13396
- tarea   1 1.8819e+12 2.9304e+13 13403
- ensuit  1 2.4572e+12 2.9879e+13 13414
- age     1 3.0936e+12 3.0515e+13 13425
- garag   1 3.2278e+12 3.0650e+13 13427

Step:  AIC=13367.98
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + bike + barb + balc + elev + fitg + party + categ

         Df  Sum of Sq        RSS   AIC
- party   1 9.9242e+10 2.7567e+13 13368
<none>                2.7468e+13 13368
- trans   1 1.1453e+11 2.7582e+13 13368
- bike    1 1.3630e+11 2.7604e+13 13369
+ bath    1 4.5750e+10 2.7422e+13 13369
- barb    1 1.9007e+11 2.7658e+13 13370
+ kidca   1 6.7460e+09 2.7461e+13 13370
+ school  1 1.1564e+09 2.7466e+13 13370
+ health  1 1.4873e+08 2.7468e+13 13370
- plaz    1 3.2229e+11 2.7790e+13 13372
- balc    1 3.6283e+11 2.7830e+13 13373
- park    1 4.0459e+11 2.7872e+13 13374
- parea   1 1.0103e+12 2.8478e+13 13386
- fitg    1 1.0114e+12 2.8479e+13 13386
- elev    1 1.0121e+12 2.8480e+13 13386
- categ   1 1.4863e+12 2.8954e+13 13394
```

```
- tarea   1 1.9670e+12 2.9435e+13 13403
- age     1 3.0670e+12 3.0535e+13 13423
- garag   1 3.4355e+12 3.0903e+13 13430
- ensuit  1 3.9879e+12 3.1456e+13 13439

Step:  AIC=13367.93
price ~ age + parea + tarea + ensuit + garag + plaz + park +
    trans + bike + barb + balc + elev + fitg + categ

         Df  Sum of Sq        RSS   AIC
<none>                2.7567e+13 13368
+ party   1 9.9242e+10 2.7468e+13 13368
- trans   1 1.1254e+11 2.7679e+13 13368
- bike    1 1.2648e+11 2.7693e+13 13368
+ bath    1 5.6603e+10 2.7510e+13 13369
- barb    1 1.8247e+11 2.7749e+13 13370
+ kidca   1 8.4695e+09 2.7558e+13 13370
+ school  1 1.4266e+09 2.7565e+13 13370
+ health  1 7.2368e+07 2.7567e+13 13370
- plaz    1 3.4139e+11 2.7908e+13 13373
- park    1 3.9836e+11 2.7965e+13 13374
- balc    1 4.9817e+11 2.8065e+13 13376
- elev    1 9.1485e+11 2.8482e+13 13384
- parea   1 1.0117e+12 2.8579e+13 13385
- fitg    1 1.5360e+12 2.9103e+13 13395
- categ   1 1.5970e+12 2.9164e+13 13396
- tarea   1 2.0311e+12 2.9598e+13 13404
- age     1 3.0441e+12 3.0611e+13 13423
- garag   1 3.4395e+12 3.1006e+13 13430
- ensuit  1 3.9841e+12 3.1551e+13 13439
Call:
lm(formula = price ~ age + parea + tarea + ensuit + garag + plaz +
    park + trans + bike + barb + balc + elev + fitg + categ,
    data = imoveiscwbav)

Coefficients:
(Intercept)          age        parea        tarea       ensuit        garag
    -432184        -7608         2674         2048       135392       172755
       plaz         park        trans         bike         barb         balc
     227540       -62911        29844       -81354       -41485        74630
       elev         fitg        categ
     -98464       138618       295414
```

Portanto o melhor modelo é:

In [230…
```r
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
-498873 -131074   -2422  106907 2398847

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -432184.4   105985.0  -4.078 5.25e-05 ***
age           -7608.4      998.3  -7.621 1.18e-13 ***
parea          2674.3      608.7   4.394 1.35e-05 ***
tarea          2047.9      329.0   6.225 9.83e-10 ***
ensuit       135391.5    15528.3   8.719  < 2e-16 ***
garag        172754.5    21324.7   8.101 3.83e-15 ***
plaz         227540.0    89152.2   2.552  0.01098 *
park         -62910.6    22818.6  -2.757  0.00604 **
trans         29843.6    20366.1   1.465  0.14342
bike         -81354.4    52369.0  -1.553  0.12091
barb         -41484.8    22232.6  -1.866  0.06261 .
balc          74630.0    24206.1   3.083  0.00216 **
elev         -98463.8    23566.9  -4.178 3.44e-05 ***
fitg         138618.1    25605.3   5.414 9.40e-08 ***
categ        295414.3    53516.0   5.520 5.33e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 228900 on 526 degrees of freedom
Multiple R-squared:  0.8089,    Adjusted R-squared:  0.8038
F-statistic:   159 on 14 and 526 DF,  p-value: < 2.2e-16
```

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

In [231…
```r
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 = 290, df = 14, p-value < 2.2e-16
```

Obtendo o valor quiquadrado tabelado

In [232…
```r
qchisq(0.95, df=14, lower.tail = TRUE)
```

23.6847913048406

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

In [234…
```r
library(robust)

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

```
00:00:01 left
00:00:01 left
00:00:01 left
00:00:01 left
00:00:01 left
00:00:01 left
00:00:01 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
```

```
Warning message in lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control, :
"Max iteration for refinement reached."
```

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

Residuals:
    Min      1Q  Median      3Q     Max
-479369  -97422    4898  115760 2507441

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -483934.3    83244.9  -5.813 1.06e-08 ***
age           -7293.5      741.1  -9.841  < 2e-16 ***
parea          2681.1      483.5   5.545 4.66e-08 ***
tarea          1549.4      284.9   5.439 8.22e-08 ***
ensuit       134199.9    13209.0  10.160  < 2e-16 ***
garag        146425.7    16707.0   8.764  < 2e-16 ***
plaz         321773.4    67936.3   4.736 2.80e-06 ***
park         -32549.7    17591.4  -1.850  0.06483 .
trans         36201.5    15078.7   2.401  0.01670 *
bike         -97602.6    40426.2  -2.414  0.01610 *
barb          -7928.0    17619.6  -0.450  0.65293
balc          54352.5    18933.1   2.871  0.00426 **
elev         -64990.3    19230.2  -3.380  0.00078 ***
fitg         100668.5    20340.2   4.949 1.00e-06 ***
categ        371335.4    41608.0   8.925  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 144000 on 526 degrees of freedom
Multiple R-Squared: 0.6073

Test for Bias:
            statistic   p-value
M-estimate      51.49 6.868e-06
LS-estimate     36.60 1.448e-03
```

h) Obtenha os indicadores de desempenho do modelo;

In [235…

```
AIC(resultados)
BIC(resultados)
```

14905.2230711426

14973.9177796041

In [236…
```
library(AICcmodavg)
AICc(resultados)
```

14906.2612390815

In [173…
```
library(performance)
model_performance(resultados)
model_performance(resultrob)
```

AIC	BIC	R2	R2_adjusted	RMSE	Sigma
14905.22	14973.92	0.8088607	0.8037734	225733.1	228929.1

R2	RMSE	Sigma
0.7905655	236289.5	239635

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

In [237…
```
# Calculando os intervalos de confiança - Para regressão linear "normal"

confint(resultados, level = 0.95)
```

	2.5 %	97.5 %
(Intercept)	-640390.197	-223978.663
age	-9569.572	-5647.220
parea	1478.579	3869.970
tarea	1401.687	2694.174
ensuit	104886.352	165896.726
garag	130862.467	214646.556
plaz	52401.923	402678.148
park	-107737.383	-18083.784
trans	-10165.240	69852.514
bike	-184232.529	21523.711
barb	-85160.380	2190.798
balc	27077.574	122182.509
elev	-144760.587	-52166.944
fitg	88316.975	188919.288
categ	190282.870	400545.700

In [238…

```
# Calculando os intervalos de confiança - Para a regressão robusta

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.

In [239…
```
#age
log(5)
```

1.6094379124341

In [240…
```
#educ
log(150)
```

5.01063529409626

In [241…
```
#earns
```

```
log(190)
```

5.24702407216049

In [242…
```
# plaz
log(0.08)
```

-2.52572864430826

In [243…
```
#parea
log(150)
```

5.01063529409626

In [244…
```
#tarea
log(190)
```

5.24702407216049

In [249…
```
#kidlt6 = 1 --> tem filhos com menos de 6 anos
#kidge6 =0 --> parâmetro não significativo a 95%
#union = 0  --> parâmetro não significativo a 95%
#husearns = 0 --> parâmetro não significativo a 95%
#hushrs = 0 --> parâmetro não significativo

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: -397509.870811602

In [255…
```
exp(-397509.870811602)
```

0

In [256…
```
anova.lmRob(resultrob)
anova(resultados)
```

```
Warning message in lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control, :
"Max iteration for refinement reached."Warning message in lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control =
robust.control, :
"Max iteration for refinement reached."
```

```
00:00:00 left
```
```
Warning message in lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control, :
"Max iteration for refinement reached."
```
```
00:00:00 left
00:00:00 left
00:00:00 left
```
```
Warning message in lmRob.fit.compute(x, y, x1.idx = x1.idx, nrep = nrep, robust.control = robust.control, :
"Max iteration for refinement reached."
```
```
00:00:01 left
00:00:01 left
00:00:01 left
00:00:00 left
00:00:00 left
00:00:00 left
00:00:00 left
```

	Chisq	Df	RobustF	Pr(F)
(Intercept)	1		NA	NA
age	1		205.08139660	0.000000e+00
parea	1		410.02790435	0.000000e+00
tarea	1		84.91062412	0.000000e+00
ensuit	1		77.36769245	0.000000e+00
garag	1		58.07340719	8.104628e-15
plaz	1		11.58140111	5.245293e-04
park	1		14.02048281	1.358258e-04
trans	1		1.72004165	1.813986e-01
bike	1		12.32212109	3.474436e-04
barb	1		0.06918982	7.886646e-01
balc	1		5.49291547	1.692666e-02
elev	1		5.11284106	2.121207e-02
fitg	1		28.85005407	4.413883e-08
categ	1		55.34172989	3.430589e-14

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	4.118120e+13	4.118120e+13	785.7727779	1.881083e-106
parea	1	4.962541e+13	4.962541e+13	946.8954239	1.063549e-119
tarea	1	7.586388e+12	7.586388e+12	144.7548093	1.271275e-29
ensuit	1	6.588508e+12	6.588508e+12	125.7144024	2.614428e-26
garag	1	4.407336e+12	4.407336e+12	84.0957552	1.062007e-18
plaz	1	7.644615e+10	7.644615e+10	1.4586583	2.276865e-01
park	1	2.016816e+12	2.016816e+12	38.4825855	1.119422e-09
trans	1	2.261392e+10	2.261392e+10	0.4314930	5.115445e-01
bike	1	2.333090e+11	2.333090e+11	4.4517355	3.533747e-02
barb	1	3.777160e+10	3.777160e+10	0.7207145	3.962958e-01
balc	1	9.103650e+11	9.103650e+11	17.3705471	3.595692e-05
elev	1	3.750541e+11	3.750541e+11	7.1563540	7.702535e-03
fitg	1	1.999045e+12	1.999045e+12	38.1434931	1.316225e-09
categ	1	1.596971e+12	1.596971e+12	30.4715859	5.327990e-08
Residuals	526	2.756689e+13	5.240854e+10	NA	NA

Nota:

Resultados foram Heterogêneos e inconsistentes. Infelizmente não fui capaz de tratar as informações para atingir um resultado satisfatório. Estou estudando para ver se minha dificuldade está nos dados, nos métodos ou algum conflito de bibliotecas das ferramentas que estou usando.

In [ ]: