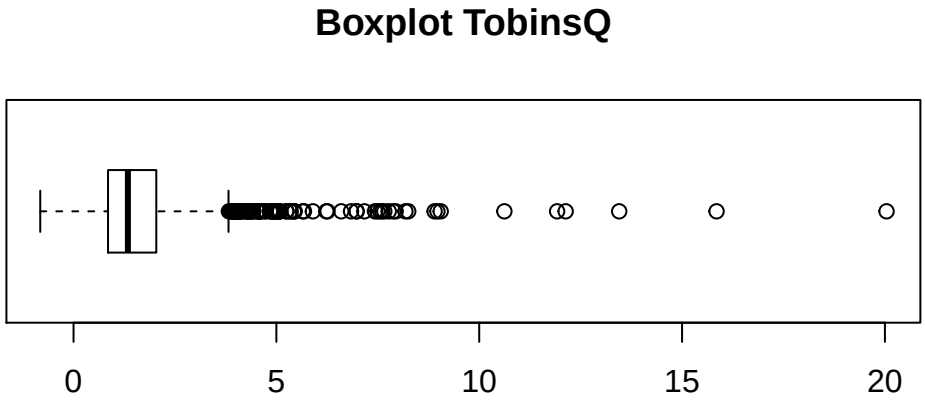
Results

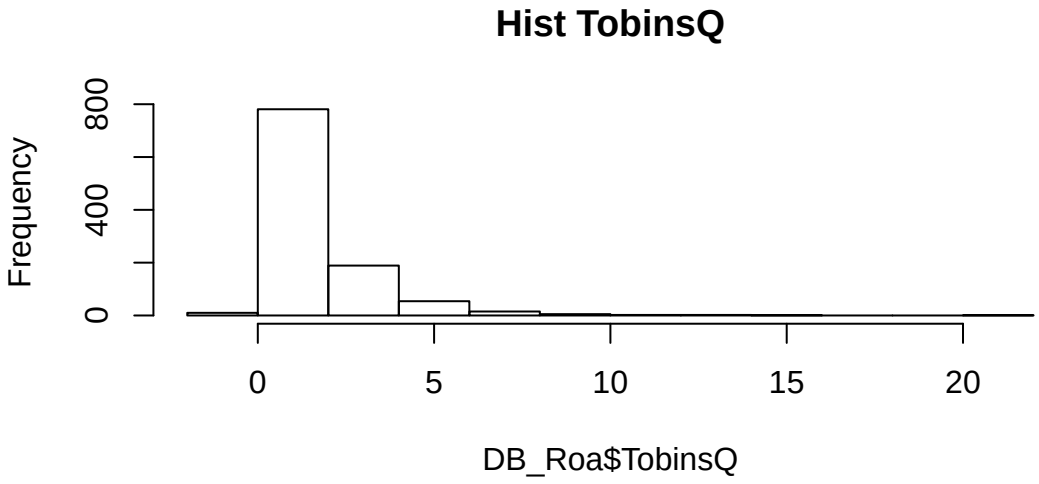
Descriptive Statistics

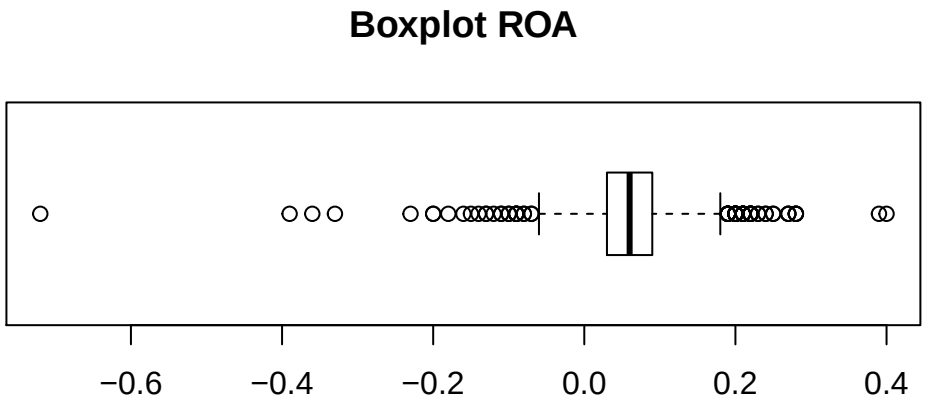
Table 1: Statistic Descriptive

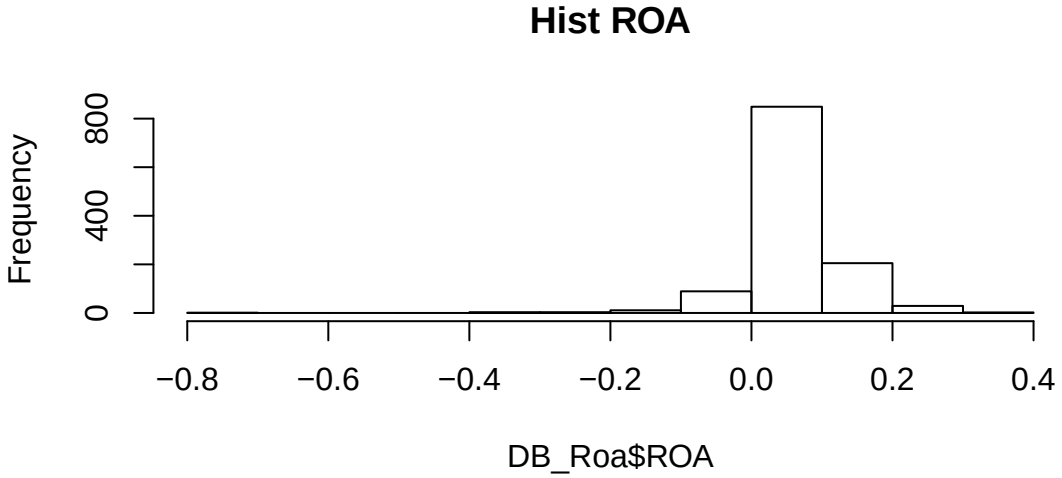
Statistic	N	Mean	St. Dev.	Min	Max
ROA	1,192	0.062	0.070	−0.720	0.400
TobinsQ	1,060	1.738	1.679	−0.820	20.040
EnergyProductivity	1,192	0.110	0.200	0.000	0.970
CarbonProductivity	1,192	0.119	0.182	0.000	0.970
WaterProductivity	1,192	0.085	0.184	0.000	0.990
WasteProductivity	1,192	0.072	0.169	0.000	0.970
SustainabilityPayLink	1,192	0.049	0.050	0.000	0.100
SustainableThemedCommitment	1,192	0.024	0.025	0.000	0.050
AuditScore	1,192	0.023	0.025	0.000	0.050
Industry	1,192	4.589	2.666	1	11
FirmSize	1,192	10.374	0.598	8.740	13.910
Leverage	1,192	3.798	42.377	0.000	875.590
NetMargin	1,192	0.105	0.220	−3.600	1.630

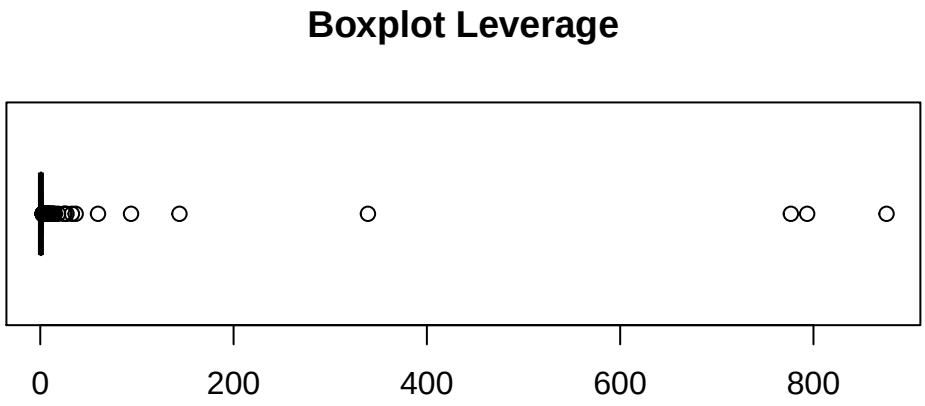
Some boxplots and histogram

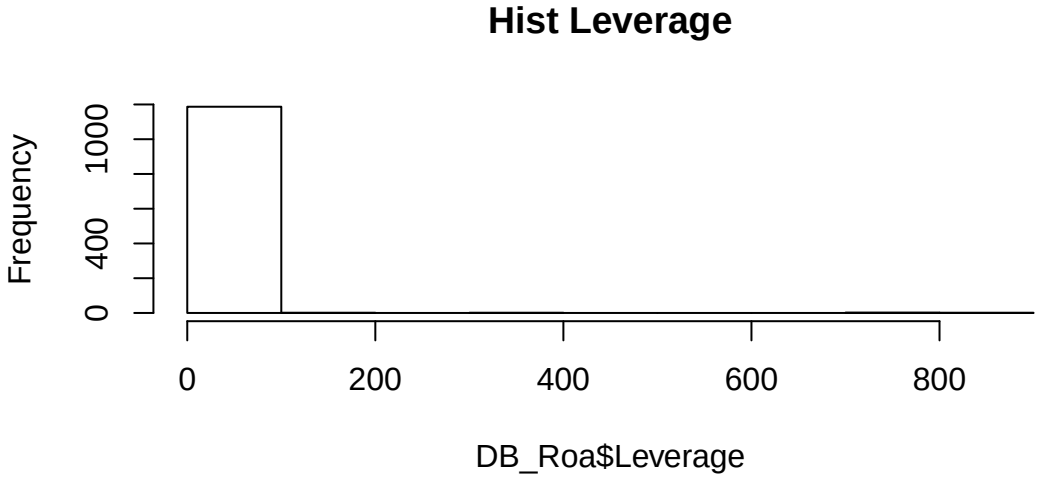


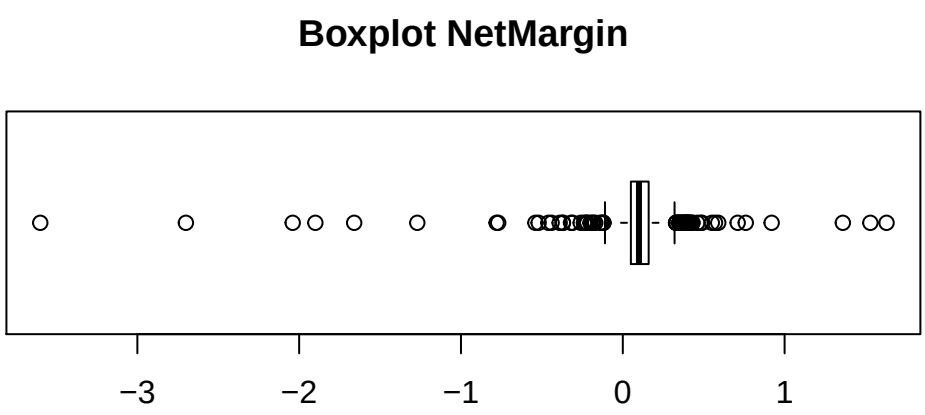


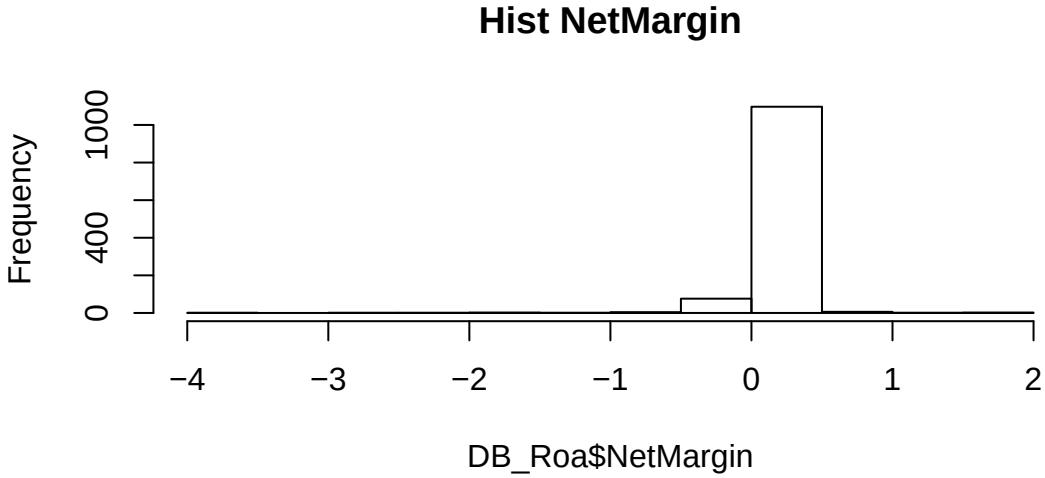

Boxplot ROA
−0.6
−0.4
−0.2
0.0
0.2
0.4

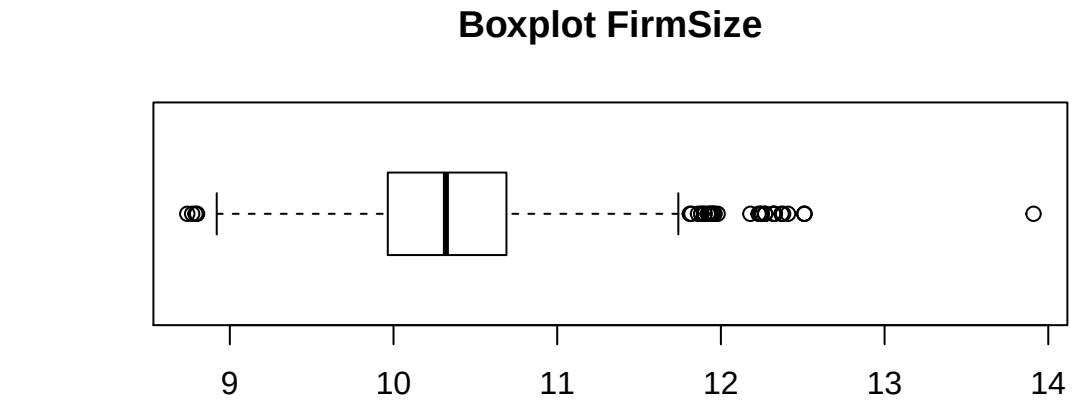

Hist ROA
Frequency
0
400
800
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
DB_Roa$ROA

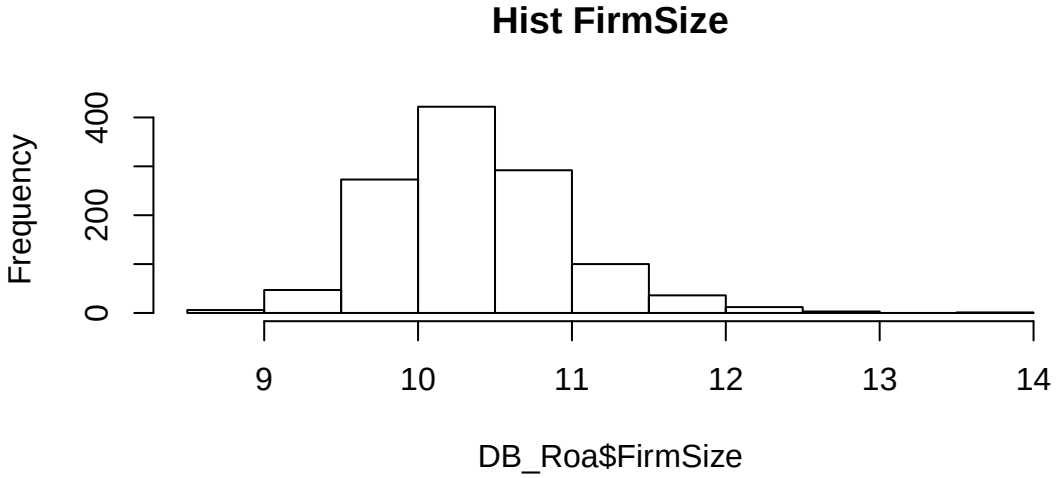

Boxplot Leverage
0
200
400
600
800


Hist Leverage
Frequency
0
400
1000
0
200
400
600
800
DB_Roa$Leverage


Boxplot NetMargin
−3
−2
−1
0
1


Hist NetMargin
Frequency
0
400
1000
−4
−3
−2
−1
0
1
2
DB_Roa$NetMargin

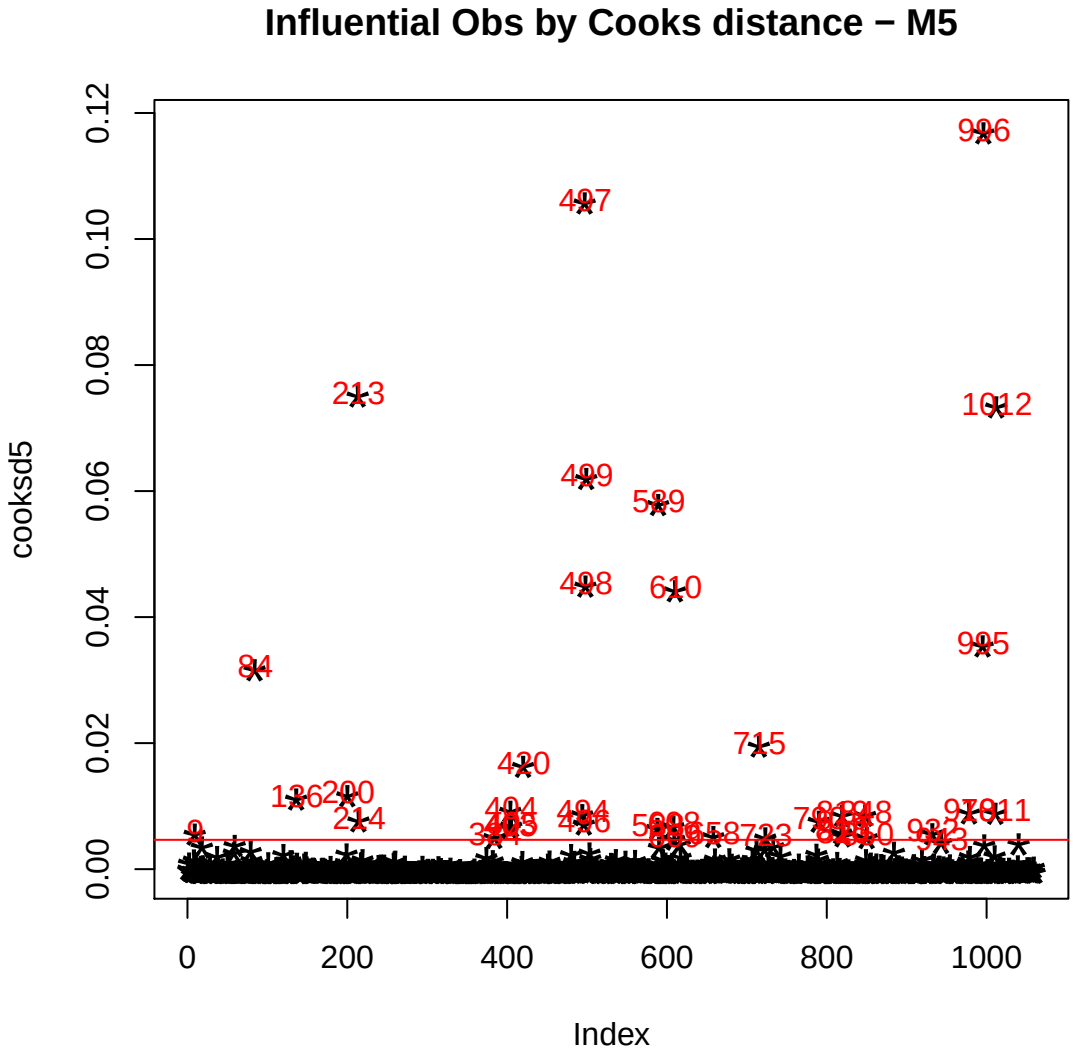

Boxplot FirmSize
9
10
11
12
13
14

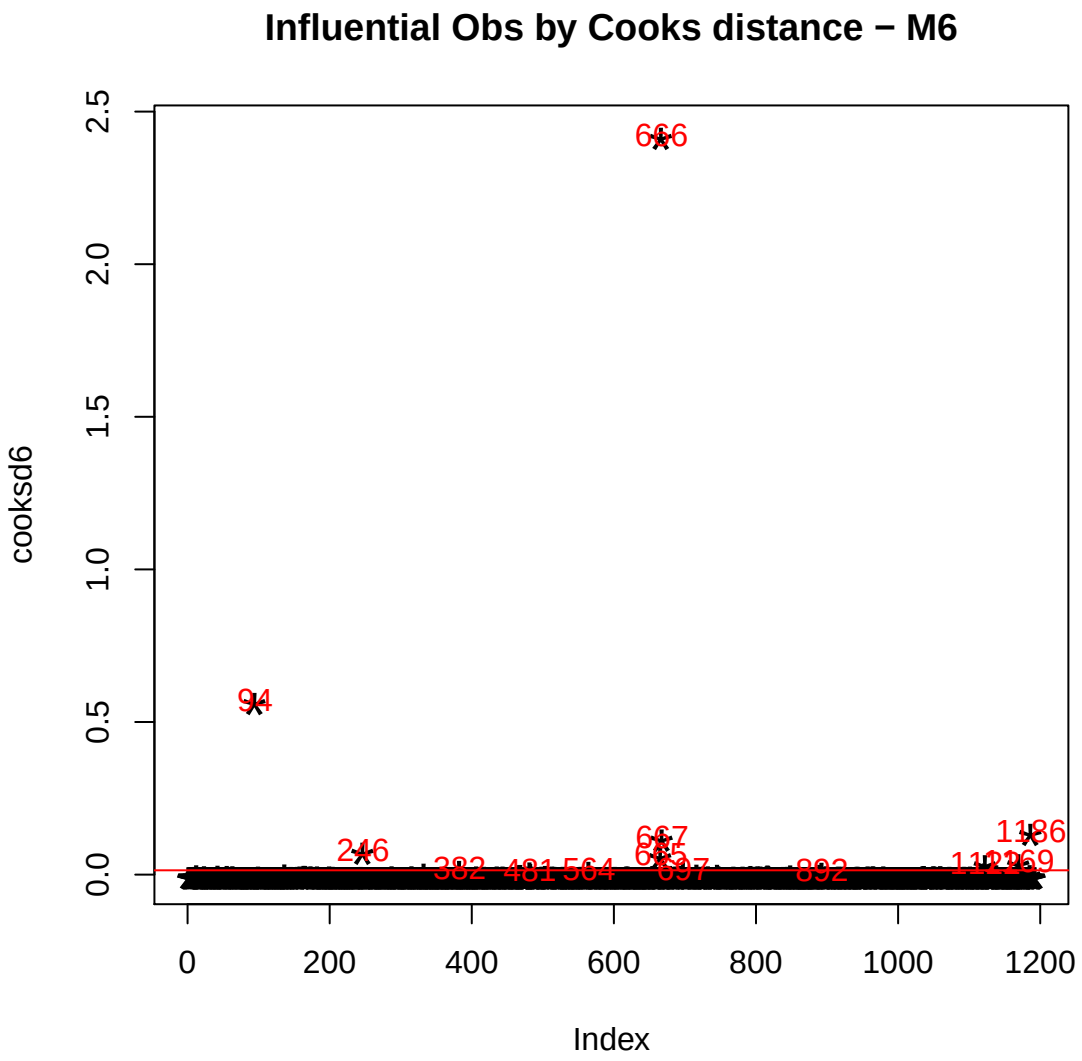

Hist FirmSize
Frequency
0
200
400
9
10
11
12
13
14
DB_Roa$FirmSize

Cooks Distance

This section will not be in the final document. Here I measure the cook's distance of my model 5 and 6. Cook's distance is a measure computed with respect to a given regression model and therefore is impacted only by the X variables included in the model. Cook's distance computes the influence exerted by each data point (row) on the predicted outcome. I summarise on a graph (i.e. one for each model), those observations that have a cook's distance greater than 4 times the mean and which may be classified as influential. I want to detect which observations is an outlier. See below both graphics.

Should I redo this process for each Model?





Here the function outlierTest from *car* package gives the most extreme observation based on the given model.
Should I remove those observations from my database?

```
## The function outlierTest from car package gives the most extreme observation based on the given model

car::outlierTest(M6)
```

```
##       rstudent unadjusted p-value Bonferonni p
## 666   9.905081         2.8491e-22   3.3961e-19
## 382  -8.232273         4.8257e-16   5.7522e-13
## 1186 -6.474909         1.3892e-10   1.6559e-07
## 94   -6.465151         1.4786e-10   1.7625e-07
## 481   5.119119         3.5819e-07   4.2696e-04
## 482   5.020838         5.9372e-07   7.0771e-04
## 564  -4.383331         1.2730e-05   1.5174e-02
## 848  -4.379382         1.2959e-05   1.5447e-02
```

```r
car::outlierTest(M5)
```

```
##       rstudent unadjusted p-value Bonferonni p
## 497  12.101376         1.2047e-31   1.2769e-28
## 499   8.633204         2.1908e-17   2.3222e-14
## 498   6.669377         4.1518e-11   4.4009e-08
## 995   6.227579         6.8476e-10   7.2585e-07
## 1012  6.138307         1.1819e-09   1.2528e-06
## 996   5.989238         2.8948e-09   3.0685e-06
## 819   5.441583         6.5755e-08   6.9700e-05
## 610   4.996946         6.8225e-07   7.2319e-04
## 820   4.387600         1.2620e-05   1.3377e-02
## 818   4.097120         4.5062e-05   4.7765e-02
```