

## Lab7\_Assignment

Diego Castro

11/2/2021

```
setwd ("C:/Users/usuario/OneDrive - University of East Anglia/PhD/First  
Semestre/Econometrics/Laboratories/Lab7") list.files()
```

```
install.packages("MASS") install.packages("GGally") install.packages("Metrics")
```

```
load("C:/Users/usuario/OneDrive - University of East Anglia/PhD/First Semestr  
e/Econometrics/Laboratories/Lab7/car.test.RData")  
load("C:/Users/usuario/OneDrive - University of East Anglia/PhD/First Semestr  
e/Econometrics/Laboratories/Lab7/car.train.RData")
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.1.1
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':  
##   method from  
##   +.gg      ggplot2
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.1.1
```

```
library (ggplot2)  
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.1.1
```

```
library(Metrics)
```

```
## Warning: package 'Metrics' was built under R version 4.1.1
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 4.1.1
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary St  
atistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

#1. a) Explore the dataset and obtain descriptive statistics.

```
dt.carttest <- data.table(car.test)
rm(car.test)
```

```
stargazer(dt.carttest, type = "text")
```

```
##
## =====
## Statistic   N      Mean      St. Dev.   Min   Pctl(25) Pctl(75)   Max
## -----
## Price       574 10,770.230 3,793.344  4,350   8,495    11,950    32,500
## Age         574   56.051    18.722    1       44       70        80
## KM          574 67,780.140 37,231.270  1  42,867.2 86,081.8 217,764
## HP          574  101.204    15.554    69       86       110       192
## MetColor    574   0.678      0.468     0         0         1         1
## Automatic   574   0.051      0.219     0         0         0         1
## CC          574 1,569.721   190.699   1,300   1,400    1,600    2,000
## Doors       574   4.003      0.954     3         3         5         5
## Weight      574 1,074.014   55.684   1,000   1,040    1,100    1,480
## -----
```

```
dt.cartrain <- data.table(car.train)
rm(car.train)
```

```
stargazer(dt.cartrain, type = "text")
```

```
##
## =====
## Statistic   N      Mean      St. Dev.   Min   Pctl(25) Pctl(75)   Max
## -----
## Price       862 10,704.590 3,513.801  4,400   8,450    11,950   31,275
## Age         862   55.878   18.529    1      44      70      80
## KM          862 69,034.760 37,701.790  1    43,000  88,171.2 243,000
## HP          862  101.701   14.593    69     90     110     192
## MetColor    862   0.673    0.469     0      0       1       1
## Automatic   862   0.059    0.236     0      0       0       1
## CC          862  1,564.901 184.890   1,300   1,400   1,600   2,000
## Doors       862   4.053    0.952     2      3       5       5
## Weight      862  1,071.425  50.520   1,000   1,040   1,085   1,615
## -----
```

#1. b) How good is Age at predicting Price?

```
lm.cartrainprice <- lm(Price~ Age, data = dt.cartrain)
stargazer(lm.cartrainprice, type = "text")
```

```
##
## =====
##               Dependent variable:
##               -----
```

```
##                               Price
## -----
## Age                           -167.361***
##                               (3.041)
##
## Constant                      20,056.420***
##                               (179.011)
##
## -----
## Observations                   862
## R2                             0.779
## Adjusted R2                   0.779
## Residual Std. Error    1,653.344 (df = 860)
## F Statistic             3,028.937*** (df = 1; 860)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

*# Answer R2 of 0.79*

*#1. c) Use the function step to improve your prediction model.*

```
step.model <- stepAIC(lm.cartrainprice, direction = "both", trace = FALSE)
summary(step(lm.cartrainprice))
```

```
## Start:  AIC=12777.8
## Price ~ Age
##
##           Df Sum of Sq      RSS   AIC
## <none>                2.3509e+09 12778
## - Age      1  8279744523 1.0631e+10 14076
##
## Call:
## lm(formula = Price ~ Age, data = dt.cartrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6234.2  -942.0    68.6   832.5 11888.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20056.425    179.011   112.04  <2e-16 ***
## Age         -167.361     3.041   -55.04  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1653 on 860 degrees of freedom
## Multiple R-squared:  0.7789, Adjusted R-squared:  0.7786
## F-statistic:  3029 on 1 and 860 DF,  p-value: < 2.2e-16
```

```
lm.cartrainprice2 <- lm(Price~ . , data = dt.cartrain)

summary(step(lm.cartrainprice2))

## Start:  AIC=12328.75
## Price ~ Age + KM + FuelType + HP + MetColor + Automatic + CC +
##      Doors + Weight
##
##              Df  Sum of Sq      RSS   AIC
## - Doors        1    179666 1367641272 12327
## - MetColor      1    474439 1367936045 12327
## <none>                                1367461607 12329
## - Automatic    1    8691751 1376153358 12332
## - FuelType     2    44312263 1411773870 12352
## - CC           1    64789102 1432250709 12367
## - HP           1   116727742 1484189348 12397
## - KM           1   139399741 1506861348 12410
## - Weight       1   216426919 1583888525 12453
## - Age          1  2310948919 3678410526 13180
##
## Step:  AIC=12326.86
## Price ~ Age + KM + FuelType + HP + MetColor + Automatic + CC +
##      Weight
##
##              Df  Sum of Sq      RSS   AIC
## - MetColor      1    501801 1368143073 12325
## <none>                                1367641272 12327
## - Automatic    1    8512151 1376153424 12330
## - FuelType     2    45244655 1412885928 12351
## - CC           1    65803225 1433444498 12365
## - HP           1   120648969 1488290242 12398
## - KM           1   139342595 1506983868 12408
## - Weight       1   255983844 1623625116 12473
## - Age          1  2315709821 3683351093 13179
##
## Step:  AIC=12325.17
## Price ~ Age + KM + FuelType + HP + Automatic + CC + Weight
##
##              Df  Sum of Sq      RSS   AIC
## <none>                                1368143073 12325
## - Automatic    1    8278648 1376421721 12328
## - FuelType     2    44943988 1413087060 12349
## - CC           1    65368607 1433511679 12363
## - HP           1   120420014 1488563087 12396
## - KM           1   140102634 1508245706 12407
## - Weight       1   256991562 1625134635 12472
## - Age          1  2318274423 3686417496 13178
##
## Call:
```

```
## lm(formula = Price ~ Age + KM + FuelType + HP + Automatic + CC +
##      Weight, data = dt.cartrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9774.4  -740.3    14.4    715.0   6571.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.097e+03  1.601e+03  -1.310   0.1904
## Age          -1.233e+02  3.243e+00 -38.018 < 2e-16 ***
## KM           -1.516e-02  1.622e-03  -9.346 < 2e-16 ***
## FuelTypeDiesel 3.563e+03  6.819e+02   5.224 2.20e-07 ***
## FuelTypePetrol 1.154e+03  4.592e+02   2.513  0.0122 *
## HP            6.324e+01  7.299e+00   8.665 < 2e-16 ***
## Automatic     4.294e+02  1.890e+02   2.272  0.0233 *
## CC           -4.380e+00  6.861e-01  -6.384 2.83e-10 ***
## Weight        1.843e+01  1.456e+00  12.658 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1266 on 853 degrees of freedom
## Multiple R-squared:  0.8713, Adjusted R-squared:  0.8701
## F-statistic: 721.9 on 8 and 853 DF,  p-value: < 2.2e-16

stargazer(lm.cartrainprice2, type = 'text')

##
## =====
##                               Dependent variable:
##                               -----
##                               Price
##                               -----
## Age                           -123.229***
##                               (3.249)
##
## KM                            -0.015***
##                               (0.002)
##
## FuelTypeDiesel                3,618.665***
##                               (695.323)
##
## FuelTypePetrol                1,140.644**
##                               (460.734)
##
## HP                            63.870***
##                               (7.494)
##
## MetColor                      50.602
##                               (93.126)
```

```

##
## Automatic          445.368**
##                   (191.495)
##
## CC                -4.445***
##                   (0.700)
##
## Doors             17.076
##                   (51.069)
##
## Weight            18.213***
##                   (1.569)
##
## Constant          -1,925.725
##                   (1,668.404)
##
## -----
## Observations      862
## R2                 0.871
## Adjusted R2       0.870
## Residual Std. Error 1,267.631 (df = 851)
## F Statistic       576.464*** (df = 10; 851)
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01

lm.cartrainprice3 <- lm(Price~ Age + KM + FuelType + HP + MetColor + Automati
c + CC + Doors + Weight, data = dt.cartrain)
stargazer(lm.cartrainprice3, type = 'text')

##
## =====
##                   Dependent variable:
##                   -----
##                   Price
## -----
## Age                -123.229***
##                   (3.249)
##
## KM                 -0.015***
##                   (0.002)
##
## FuelTypeDiesel     3,618.665***
##                   (695.323)
##
## FuelTypePetrol     1,140.644**
##                   (460.734)
##
## HP                 63.870***
##                   (7.494)
##

```

```

## MetColor                50.602
##                        (93.126)
##
## Automatic                445.368**
##                        (191.495)
##
## CC                       -4.445***
##                        (0.700)
##
## Doors                    17.076
##                        (51.069)
##
## Weight                   18.213***
##                        (1.569)
##
## Constant                 -1,925.725
##                        (1,668.404)
##
## -----
## Observations              862
## R2                        0.871
## Adjusted R2               0.870
## Residual Std. Error      1,267.631 (df = 851)
## F Statistic               576.464*** (df = 10; 851)
## =====
## Note:                     *p<0.1; **p<0.05; ***p<0.01

lm.cartrainprice4 <- lm(Price~ Age + KM + FuelType + HP + Automatic + CC + Weight, data = dt.cartrain)
stargazer(lm.cartrainprice4, type = 'text')

##
## =====
##                        Dependent variable:
##                        -----
##                        Price
## -----
## Age                      -123.309***
##                        (3.243)
##
## KM                       -0.015***
##                        (0.002)
##
## FuelTypeDiesel           3,562.514***
##                        (681.934)
##
## FuelTypePetrol           1,153.977**
##                        (459.248)
##
## HP                       63.245***

```

```
## (7.299)
##
## Automatic 429.382**
## (188.997)
##
## CC -4.380***
## (0.686)
##
## Weight 18.433***
## (1.456)
##
## Constant -2,097.400
## (1,600.659)
##
## -----
## Observations 862
## R2 0.871
## Adjusted R2 0.870
## Residual Std. Error 1,266.459 (df = 853)
## F Statistic 721.861*** (df = 8; 853)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01

dt.cartrain$age2 <- (dt.cartrain$Age)*(dt.cartrain$Age)
dt.cartrain
```

```
## Price Age KM FuelType HP MetColor Automatic CC Doors Weight age
2
## 1: 13750 23 72937 Diesel 90 1 0 2000 3 1165 52
9
## 2: 13950 24 41711 Diesel 90 1 0 2000 3 1165 57
6
## 3: 13750 30 38500 Diesel 90 0 0 2000 3 1170 90
0
## 4: 12950 32 61000 Diesel 90 0 0 2000 3 1170 102
4
## 5: 16900 27 94612 Diesel 90 1 0 2000 3 1245 72
9
## ---
## 858: 8950 71 29000 Petrol 86 1 1 1300 3 1045 504
1
## 859: 8450 72 26000 Petrol 86 0 0 1300 3 1015 518
4
## 860: 7500 69 20544 Petrol 86 1 0 1300 3 1025 476
1
## 861: 8500 71 17016 Petrol 86 0 0 1300 3 1015 504
1
## 862: 6950 76 1 Petrol 110 0 0 1600 5 1114 577
6
```



```
## =====
##                               Dependent variable:
##                               -----
##                               Price
## -----
## Age                           -251.416***
##                               (11.679)
##
## KM                             -0.014***
##                               (0.002)
##
## FuelTypeDiesel                3,737.256***
##                               (636.086)
##
## FuelTypePetrol                865.711**
##                               (428.998)
##
## HP                             68.304***
##                               (6.821)
##
## Automatic                     570.180***
##                               (176.674)
##
## CC                             -4.163***
##                               (0.640)
##
## Weight                        13.244***
##                               (1.433)
##
## age2                          1.264***
##                               (0.111)
##
## Constant                      5,548.105***
##                               (1,637.408)
## -----
## Observations                   862
## R2                             0.888
## Adjusted R2                   0.887
## Residual Std. Error    1,180.967 (df = 852)
## F Statistic            752.248*** (df = 9; 852)
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

```

dt.cartrain$KM2 <- (dt.cartrain$KM)^2
dt.cartrain$KM_KM <- (dt.cartrain$KM)*(dt.cartrain$KM)

## Warning in (dt.cartrain$KM) * (dt.cartrain$KM): NAs produced by integer ov
erflow

lm.cartrainprice6 <- lm(Price~ Age + KM + FuelType + HP + Automatic + CC + We
ight + age2 + KM2, data = dt.cartrain)
stargazer(lm.cartrainprice6, type = 'text')

##
## =====
##                               Dependent variable:
##                               -----
##                               Price
## -----
## Age                           -257.225***
##                               (12.287)
##
## KM                             -0.009**
##                               (0.004)
##
## FuelTypeDiesel                 3,745.599***
##                               (635.632)
##
## FuelTypePetrol                 849.846**
##                               (428.804)
##
## HP                             67.762***
##                               (6.825)
##
## Automatic                     582.161***
##                               (176.720)
##
## CC                             -4.129***
##                               (0.640)
##
## Weight                        13.113***
##                               (1.434)
##
## age2                          1.306***
##                               (0.115)
##
## KM2                           -0.00000
##                               (0.00000)
##
## Constant                      5,686.335***
##                               (1,638.734)
## -----

```

```
## Observations      862
## R2                0.889
## Adjusted R2       0.887
## Residual Std. Error 1,180.079 (df = 851)
## F Statistic      678.271*** (df = 10; 851)
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01

dt.cartrain$lprice <- log(dt.cartrain$Price)

lm.cartrainprice7 <- lm(lprice~ Age + KM + FuelType + HP + Automatic + CC + W
eight + age2 + KM2, data = dt.cartrain)
stargazer(lm.cartrainprice7, type = 'text')

##
## =====
##                               Dependent variable:
##                               -----
##                               lprice
## -----
## Age                          -0.013***
##                               (0.001)
##
## KM                           -0.00000
##                               (0.00000)
##
## FuelTypeDiesel               0.122**
##                               (0.060)
##
## FuelTypePetrol               0.072*
##                               (0.040)
##
## HP                           0.003***
##                               (0.001)
##
## Automatic                    0.050***
##                               (0.017)
##
## CC                           -0.0001
##                               (0.0001)
##
## Weight                       0.001***
##                               (0.0001)
##
## age2                         0.00002**
##                               (0.00001)
##
## KM2                          -0.000***
##                               (0.000)
##
```

```

## Constant                8.673***
##                          (0.154)
##
## -----
## Observations              862
## R2                       0.856
## Adjusted R2              0.855
## Residual Std. Error      0.111 (df = 851)
## F Statistic              507.116*** (df = 10; 851)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01

dt.cartrain$lWeight <- log(dt.cartrain$Weight)

lm.cartrainprice8 <- lm(lprice~ Age + KM + FuelType + HP + Automatic + CC + W
eight + age2 + KM2 +lWeight, data = dt.cartrain)
stargazer(lm.cartrainprice8, type = 'text')

##
## =====
##                               Dependent variable:
##                               -----
##                               lprice
## -----
## Age                         -0.012***
##                             (0.001)
##
## KM                          -0.00000
##                             (0.00000)
##
## FuelTypeDiesel              0.150**
##                             (0.059)
##
## FuelTypePetrol              0.085**
##                             (0.040)
##
## HP                          0.003***
##                             (0.001)
##
## Automatic                   0.034**
##                             (0.017)
##
## CC                          -0.0002***
##                             (0.0001)
##
## Weight                      -0.005***
##                             (0.001)
##
## age2                        0.00001
##                             (0.00001)

```

```
##
## KM2 -0.000***
## (0.000)
##
## lWeight 7.070***
## (1.425)
##
## Constant -34.251***
## (8.653)
##
## -----
## Observations 862
## R2 0.860
## Adjusted R2 0.859
## Residual Std. Error 0.109 (df = 850)
## F Statistic 476.044*** (df = 11; 850)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01

dt.cartrain$HP2 <- (dt.cartrain$HP)*(dt.cartrain$HP)

lm.cartrainprice9 <- lm(lprice~ Age + KM + FuelType + HP + Automatic + CC + W
eight + age2 + KM2 +lWeight + HP2, data = dt.cartrain)
stargazer(lm.cartrainprice9, type = 'text')

##
## =====
## Dependent variable:
## -----
## lprice
## -----
## Age -0.012***
## (0.001)
##
## KM -0.00000
## (0.00000)
##
## FuelTypeDiesel 0.069
## (0.080)
##
## FuelTypePetrol 0.082**
## (0.040)
##
## HP -0.001
## (0.003)
##
## Automatic 0.032*
## (0.017)
##
## CC -0.0001
```

```

##                                (0.0001)
##
## Weight                        -0.005***
##                                (0.001)
##
## age2                          0.00001
##                                (0.00001)
##
## KM2                           -0.000***
##                                (0.000)
##
## lWeight                       6.726***
##                                (1.442)
##
## HP2                           0.00001
##                                (0.00001)
##
## Constant                     -32.076***
##                                (8.762)
##
## -----
## Observations                  862
## R2                           0.861
## Adjusted R2                  0.859
## Residual Std. Error         0.109 (df = 849)
## F Statistic                 437.259*** (df = 12; 849)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01

dt.cartrain$lHP <- log(dt.cartrain$HP)

lm.cartrainprice10 <- lm(lprice~ Age + FuelType + HP + Automatic + CC + Weight + age2 + KM2 + HP2 + lHP, data = dt.cartrain)
stargazer(lm.cartrainprice10, type = 'text')

##
## =====
##                                Dependent variable:
##                                -----
##                                lprice
## -----
## Age                          -0.014***
##                                (0.001)
##
## FuelTypeDiesel               -0.089
##                                (0.095)
##
## FuelTypePetrol               0.069*
##                                (0.040)
##

```

```

## HP -0.070*
## (0.039)
##
## Automatic 0.046***
## (0.017)
##
## CC 0.0002*
## (0.0001)
##
## Weight 0.001***
## (0.0001)
##
## age2 0.00003**
## (0.00001)
##
## KM2 -0.000***
## (0.000)
##
## HP2 0.0002*
## (0.0001)
##
## lHP 3.708*
## (2.168)
##
## Constant -3.151
## (6.994)
##
## -----
## Observations 862
## R2 0.858
## Adjusted R2 0.856
## Residual Std. Error 0.111 (df = 850)
## F Statistic 465.367*** (df = 11; 850)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01

lm.cartrainprice11 <- lm(lprice~ Age + FuelType + HP + Automatic + CC + Weight + age2 + KM_KM + HP2 + lHP, data = dt.cartrain)
stargazer(lm.cartrainprice11, type = 'text')

##
## =====
## Dependent variable:
## -----
## lprice
## -----
## Age -0.011***
## (0.001)
##
## FuelTypeDiesel 0.306*

```

```
## (0.158)
##
## FuelTypePetrol 0.388***
## (0.097)
##
## HP 0.260**
## (0.106)
##
## Automatic 0.024
## (0.023)
##
## CC -0.0003
## (0.0002)
##
## Weight 0.002***
## (0.0003)
##
## age2 0.00000
## (0.00002)
##
## KM_KM 0.000
## (0.000)
##
## HP2 -0.001**
## (0.0002)
##
## lHP -15.353**
## (6.295)
##
## Constant 57.900***
## (20.622)
##
## -----
## Observations 246
## R2 0.904
## Adjusted R2 0.899
## Residual Std. Error 0.094 (df = 234)
## F Statistic 199.279*** (df = 11; 234)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01
```

*#2. Did you use all the variables in the dataset to build your model? Why?  
# No, Color and Doors were not significant but I create others variables (l  
price, age2, KM2, HP2)*

*#3-Use your model to predict used car prices in the dataset carTest.RData.*

```
fit.val <- fitted.values(lm.cartrainprice10)
head(fit.val)
```



```
##          1          2          3          4          5          6
## 9.675076 9.693499 9.626540 9.582823 9.675269 9.665999
```

```
dt.cartrain$FitVal <- fitted.values(lm.cartrainprice10)
```

*#4- Use the RMSE to compare the performance of your model in carTrain.RData and carTest.RData.*

```
dt.cartrain <- dt.cartrain[, preprice:=predict(lm.cartrainprice10, newdata=dt
.cartrain)]
head (dt.cartrain)
```

```
##      Price Age      KM FuelType HP MetColor Automatic   CC Doors Weight age2
## 1: 13750  23 72937   Diesel 90      1         0 2000     3   1165  529
## 2: 13950  24 41711   Diesel 90      1         0 2000     3   1165  576
## 3: 13750  30 38500   Diesel 90      0         0 2000     3   1170  900
## 4: 12950  32 61000   Diesel 90      0         0 2000     3   1170 1024
## 5: 16900  27 94612   Diesel 90      1         0 2000     3   1245  729
## 6: 18600  30 75889   Diesel 90      1         0 2000     3   1245  900
##           KM2      KM_KM   lprice   lWeight   HP2     lHP   FitVal preprice
## 1: 5319805969      NA 9.528794 7.060476 8100 4.49981 9.675076 9.675076
## 2: 1739807521 1739807521 9.543235 7.060476 8100 4.49981 9.693499 9.693499
## 3: 1482250000 1482250000 9.528794 7.064759 8100 4.49981 9.626540 9.626540
## 4: 3721000000      NA 9.468851 7.064759 8100 4.49981 9.582823 9.582823
## 5: 8951430544      NA 9.735069 7.126891 8100 4.49981 9.675269 9.675269
## 6: 5759140321      NA 9.830917 7.126891 8100 4.49981 9.665999 9.665999
```

```
show(dt.cartrain)
```

```
##      Price Age      KM FuelType  HP MetColor Automatic   CC Doors Weight age
2
## 1: 13750  23 72937   Diesel 90      1         0 2000     3   1165  52
9
## 2: 13950  24 41711   Diesel 90      1         0 2000     3   1165  57
6
## 3: 13750  30 38500   Diesel 90      0         0 2000     3   1170  90
0
## 4: 12950  32 61000   Diesel 90      0         0 2000     3   1170 102
4
## 5: 16900  27 94612   Diesel 90      1         0 2000     3   1245  72
9
## ---
## 858:  8950  71 29000   Petrol 86      1         1 1300     3   1045 504
1
## 859:  8450  72 26000   Petrol 86      0         0 1300     3   1015 518
4
## 860:  7500  69 20544   Petrol 86      1         0 1300     3   1025 476
1
## 861:  8500  71 17016   Petrol 86      0         0 1300     3   1015 504
1
## 862:  6950  76      1   Petrol 110      0         0 1600     5   1114 577
```

```

6
##           KM2           KM_KM   lprice   lWeight   HP2       lHP   FitVal prepr
ice
## 1: 5319805969           NA 9.528794 7.060476 8100 4.499810 9.675076 9.675
076
## 2: 1739807521 1739807521 9.543235 7.060476 8100 4.499810 9.693499 9.693
499
## 3: 1482250000 1482250000 9.528794 7.064759 8100 4.499810 9.626540 9.626
540
## 4: 3721000000           NA 9.468851 7.064759 8100 4.499810 9.582823 9.582
823
## 5: 8951430544           NA 9.735069 7.126891 8100 4.499810 9.675269 9.675
269
## ---
## 858: 841000000 841000000 9.099409 6.951772 7396 4.454347 9.106223 9.106
223
## 859: 676000000 676000000 9.041922 6.922644 7396 4.454347 9.021325 9.021
325
## 860: 422055936 422055936 8.922658 6.932448 7396 4.454347 9.063926 9.063
926
## 861: 289544256 289544256 9.047821 6.922644 7396 4.454347 9.034686 9.034
686
## 862:           1           1 8.846497 7.015712 12100 4.700480 9.126543 9.126
543

rmse(dt.cartrain$lprice, dt.cartrain$yhat)

## [1] NaN

lm.cartrainpriceTest <- lm(Price~ . , data = dt.cartest)

dt.cartest <- dt.cartest[, prepricetest:=predict(lm.cartrainpriceTest, newdat
a=dt.cartest)]
head (dt.cartest)

##   Price Age    KM FuelType  HP MetColor Automatic   CC Doors Weight
## 1: 11290 49  80320   Petrol 110      1      1 1600    3   1070
## 2: 15950 19  51884   Petrol  97      1      0 1400    3   1100
## 3:  8500 80 100458   Petrol 110      0      0 1600    5   1085
## 4:  8900 67  54847   Petrol 110      0      0 1600    3   1050
## 5: 15950 28  29206   Petrol  97      1      0 1400    5   1110
## 6: 15950 30  67660   Petrol 110      1      0 1600    3   1105
##   prepricetest
## 1:   11703.048
## 2:   16315.640
## 3:    7638.298
## 4:    9286.247
## 5:   15787.871
## 6:   14818.942

rmse(dt.cartest$Price, dt.cartest$prepricetest)

```

```
## [1] 1376.47
```

*##5. Which measure - the RMSE of the train set or the RMSE of the test set - do you consider the most useful for assessing the predictive performance of a model in real world applications?*

*#Why?*

*#5Anwr: I consider that RMSE of Train is most useful in practice due to this method incorporates a "improved" fitted model of the regression that uses the Akaike criterion.*