4_exercise

May 31, 2019

In [5]:
```
options(encoding = 'UTF-8')

        #Loading all the necessary packages
        if (!require("CASdatasets")) install.packages("CASdatasets", repos = "http://cas.uqam.ca
        if (!require("tidyverse")) install.packages("tidyverse")
        if (!require("caret")) install.packages("caret")
        if (!require("plyr")) install.packages("plyr")
        if (!require("mgcv")) install.packages("mgcv")
        if (!require("gridExtra")) install.packages("gridExtra")
        if (!require("visreg")) install.packages("visreg")
        if (!require("MASS")) install.packages("MASS")

        require("CASdatasets")
        require("tidyverse")
        require("plyr")
        require("caret")
        require("mgcv")
        require("gridExtra")
        require("visreg")
        require("MASS")
```

```
Loading required package: CASdatasets
Loading required package: xts
Loading required package: zoo

Attaching package: zoo

The following objects are masked from package:base:

    as.Date, as.Date.numeric

Loading required package: sp
Loading required package: tidyverse
 Attaching packages  tidyverse 1.2.1
 ggplot2 3.1.0      purrr   0.2.5
 tibble  2.0.1      dplyr   0.7.8
 tidyr   0.8.2      stringr 1.3.1
 readr   1.3.1      forcats 0.3.0
```

```
 Conflicts  tidyverse_conflicts()
 dplyr::filter() masks stats::filter()
 dplyr::first()  masks xts::first()
 dplyr::lag()    masks stats::lag()
 dplyr::last()   masks xts::last()
Loading required package: caret
Loading required package: lattice

Attaching package: caret

The following object is masked from package:purrr:

    lift

Loading required package: plyr
------------------------------------------------------------------------------
You have loaded plyr after dplyr - this is likely to cause problems.
If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
library(plyr); library(dplyr)
------------------------------------------------------------------------------

Attaching package: plyr

The following objects are masked from package:dplyr:

    arrange, count, desc, failwith, id, mutate, rename, summarise,
    summarize

The following object is masked from package:purrr:

    compact

Loading required package: mgcv
Loading required package: nlme

Attaching package: nlme

The following object is masked from package:dplyr:

    collapse

This is mgcv 1.8-28. For overview type 'help("mgcv-package")'.
Loading required package: gridExtra

Attaching package: gridExtra

The following object is masked from package:dplyr:
```

```
    combine

Loading required package: visreg
Loading required package: MASS

Attaching package: MASS

The following object is masked from package:dplyr:

    select
```

1 Introduction

1.0.1 Load data

```
In [11]: ## If CASdatasets package can be loaded, run the following
         ## Loading the dataset
         # require("CASdatasets")
         # data("ausprivauto0405")

         ## split the data set for analysis
         #set.seed(85)
         # folds = createDataPartition(ausprivauto0405$ClaimNb, 0.5)
         #df_ausprivauto0405 = ausprivauto0405[folds[[1]], ]
         # save(df_ausprivauto0405, file="../df_ausprivauto0405.RData")

         ## If CASdatasets package can not be loaded, run the following:
         load("df_ausprivauto0405.RData") # load(file="path to the file df_ausprivauto0405")
         dataset <- df_ausprivauto0405 # store as dataset for easier re-use of previous code
```

1.0.2 Check data structure

```
In [14]: head(dataset)
```

	Exposure	VehValue	VehAge	VehBody	Gender	DrivAge	ClaimOcc	ClaimN
3	0.5694730	3.26	young cars	Utility	Female	young people	0	0
6	0.8542094	2.01	old cars	Hardtop	Male	older work. people	0	0
8	0.5557837	1.47	young cars	Hatchback	Male	oldest people	0	0
9	0.3613963	0.52	oldest cars	Hatchback	Female	working people	0	0
11	0.8542094	1.38	young cars	Hatchback	Male	young people	0	0
12	0.8542094	1.22	old cars	Hatchback	Male	older work. people	0	0

```
In [17]: str(dataset)
```

```
'data.frame':        33928 obs. of  9 variables:
 $ Exposure   : num  0.569 0.854 0.556 0.361 0.854 ...
 $ VehValue   : num  3.26 2.01 1.47 0.52 1.38 1.22 1 1.66 0.76 0.27 ...
```

```
 $ VehAge     : Factor w/ 4 levels "old cars","oldest cars",..: 3 1 3 2 3 1 3 1 1 2 ...
 $ VehBody    : Factor w/ 13 levels "Bus","Convertible",..: 13 4 5 5 5 5 5 10 5 5 ...
 $ Gender     : Factor w/ 2 levels "Female","Male": 1 2 2 1 2 2 1 2 2 1 ...
 $ DrivAge    : Factor w/ 6 levels "old people","older work. people",..: 5 2 3 4 5 2 2 3 2 5 ...
 $ ClaimOcc   : int  0 0 0 0 0 0 0 1 1 0 ...
 $ ClaimNb    : int  0 0 0 0 0 0 0 1 1 0 ...
 $ ClaimAmount: num  0 0 0 0 0 ...
```

```
In [20]: summary(dataset)
```

```
   Exposure          VehValue              VehAge
 Min.   :0.002738   Min.   : 0.000   old cars     :10050
 1st Qu.:0.219028   1st Qu.: 1.010   oldest cars  : 9409
 Median :0.443532   Median : 1.490   young cars   : 8312
 Mean   :0.468039   Mean   : 1.773   youngest cars: 6157
 3rd Qu.:0.709103   3rd Qu.: 2.140
 Max.   :0.999316   Max.   :23.590

         VehBody         Gender                   DrivAge
 Sedan        :11156   Female:19460   old people        :5381
 Hatchback    : 9544   Male  :14468   older work. people:8119
 Station wagon: 8140                  oldest people     :3285
 Utility      : 2189                  working people    :7832
 Truck        :  842                  young people      :6497
 Hardtop      :  800                  youngest people   :2814
 (Other)      : 1257
   ClaimOcc          ClaimNb         ClaimAmount
 Min.   :0.00000   Min.   :0.00000   Min.   :    0.0
 1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:    0.0
 Median :0.00000   Median :0.00000   Median :    0.0
 Mean   :0.06652   Mean   :0.07094   Mean   :  137.6
 3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:    0.0
 Max.   :1.00000   Max.   :3.00000   Max.   :55922.1
```

2 Descriptive Analysis

3 Fit a GLM for Claims Frequency

4 Fit a GLM for Claims Severity