

title

THE REGRESSORS

January 14th, 2021

Importing libraries:

```
library(dplyr)
library(ggplot2)
library(stringr)
library(gridExtra)
library(outliers)
library(PerformanceAnalytics)
library(foreach)
library(MASS)
library(e1071)
library(VGAM)
library(caret)
library(klaR)
library(arm)
library(caTools)
library(stepAIC)
library(LiblineAR)
```

Importing and manipulating the data:

```
credit <- read.csv('./data/credit.csv')
names(credit) <- tolower(names(credit))
```

Summary of the data (good credit rating):

```
summary(credit[credit$response==1,])
#>      obs.      chk_acct      duration      history
#> Min.      : 1.0      Min.      :0.000      Min.      : 4.00      Min.      :0.000
#> 1st Qu.: 244.8      1st Qu.: 1.000      1st Qu.: 12.00      1st Qu.: 2.000
#> Median : 482.5      Median : 2.000      Median : 18.00      Median : 2.000
#> Mean    : 494.0      Mean    : 1.866      Mean    : 19.21      Mean    : 2.707
#> 3rd Qu.: 747.5      3rd Qu.: 3.000      3rd Qu.: 24.00      3rd Qu.: 4.000
#> Max.    :1000.0      Max.    : 3.000      Max.    : 60.00      Max.    : 4.000
#>      new_car      used_car      furniture      radio.tv
#> Min.      :0.0000      Min.      :0.0000      Min.      :0.0000      Min.      :0.0000
#> 1st Qu.: 0.0000      1st Qu.: 0.0000      1st Qu.: 0.0000      1st Qu.: 0.0000
#> Median : 0.0000      Median : 0.0000      Median : 0.0000      Median : 0.0000
#> Mean    : 0.2071      Mean    : 0.1229      Mean    : 0.1757      Mean    : 0.3114
#> 3rd Qu.: 0.0000      3rd Qu.: 0.0000      3rd Qu.: 0.0000      3rd Qu.: 1.0000
#> Max.    : 1.0000      Max.    : 1.0000      Max.    : 1.0000      Max.    : 1.0000
#>      education      retraining      amount      sav_acct      employment
#> Min.      :0.00      Min.      :0.00      Min.      : 250      Min.      :0.00      Min.      :0.000
#> 1st Qu.: 0.00      1st Qu.: 0.00      1st Qu.: 1376      1st Qu.: 0.00      1st Qu.: 2.000
```

```

#> Median :0.00 Median :0.00 Median : 2244 Median :0.00 Median :2.000
#> Mean   :0.04 Mean   :0.09 Mean   : 2985 Mean   :1.29 Mean   :2.476
#> 3rd Qu.:0.00 3rd Qu.:0.00 3rd Qu.: 3635 3rd Qu.:3.00 3rd Qu.:4.000
#> Max.    :1.00 Max.    :1.00 Max.    :15857 Max.    :4.00 Max.    :4.000
#> install_rate male_div male_single male_mar_or_wid
#> Min.      :1.00 Min.      :0.00000 Min.      :0.0000 Min.      :0.00000
#> 1st Qu.    :2.00 1st Qu.    :0.00000 1st Qu.    :0.0000 1st Qu.    :0.00000
#> Median     :3.00 Median     :0.00000 Median     :1.0000 Median     :0.00000
#> Mean       :2.92 Mean       :0.04286 Mean       :0.5743 Mean       :0.09571
#> 3rd Qu.    :4.00 3rd Qu.    :0.00000 3rd Qu.    :1.0000 3rd Qu.    :0.00000
#> Max.       :4.00 Max.       :1.00000 Max.       :1.0000 Max.       :1.00000
#> co.applicant guarantor present_resident real_estate
#> Min.      :0.00000 Min.      :0.00 Min.      :1.000 Min.      :0.0000
#> 1st Qu.    :0.00000 1st Qu.    :0.00 1st Qu.    :2.000 1st Qu.    :0.0000
#> Median     :0.00000 Median     :0.00 Median     :3.000 Median     :0.0000
#> Mean       :0.03286 Mean       :0.06 Mean       :2.843 Mean       :0.3171
#> 3rd Qu.    :0.00000 3rd Qu.    :0.00 3rd Qu.    :4.000 3rd Qu.    :1.0000
#> Max.       :1.00000 Max.       :1.00 Max.       :4.000 Max.       :1.0000
#> prop_unkn_none age other_install rent
#> Min.      :0.0000 Min.      :19.00 Min.      :0.0000 Min.      :0.0000
#> 1st Qu.    :0.0000 1st Qu.    :27.00 1st Qu.    :0.0000 1st Qu.    :0.0000
#> Median     :0.0000 Median     :34.00 Median     :0.0000 Median     :0.0000
#> Mean       :0.1243 Mean       :36.22 Mean       :0.1571 Mean       :0.1557
#> 3rd Qu.    :0.0000 3rd Qu.    :42.25 3rd Qu.    :0.0000 3rd Qu.    :0.0000
#> Max.       :1.0000 Max.       :75.00 Max.       :1.0000 Max.       :1.0000
#> own_res num_credits job num_dependents
#> Min.      :0.0000 Min.      :1.000 Min.      :0.00 Min.      :1.000
#> 1st Qu.    :1.0000 1st Qu.    :1.000 1st Qu.    :2.00 1st Qu.    :1.000
#> Median     :1.0000 Median     :1.000 Median     :2.00 Median     :1.000
#> Mean       :0.7529 Mean       :1.424 Mean       :1.89 Mean       :1.156
#> 3rd Qu.    :1.0000 3rd Qu.    :2.000 3rd Qu.    :2.00 3rd Qu.    :1.000
#> Max.       :1.0000 Max.       :4.000 Max.       :3.00 Max.       :2.000
#> telephone foreign response
#> Min.      :0.0000 Min.      :0.00000 Min.      :1
#> 1st Qu.    :0.0000 1st Qu.    :0.00000 1st Qu.    :1
#> Median     :0.0000 Median     :0.00000 Median     :1
#> Mean       :0.4157 Mean       :0.04714 Mean     :1
#> 3rd Qu.    :1.0000 3rd Qu.    :0.00000 3rd Qu.    :1
#> Max.       :1.0000 Max.       :1.00000 Max.       :1

```

Summary of the data (not good credit rating):

```

summary(credit[credit$response==0,])
#> obs.      chk_acct      duration      history
#> Min.      : 2.0 Min.      :0.0000 Min.      : 6.00 Min.      :0.000
#> 1st Qu.    :274.8 1st Qu.    :0.0000 1st Qu.    :12.00 1st Qu.    :2.000
#> Median     :542.0 Median     :1.0000 Median     :24.00 Median     :2.000
#> Mean       :515.8 Mean       :0.9033 Mean       :24.86 Mean       :2.167
#> 3rd Qu.    :755.2 3rd Qu.    :1.0000 3rd Qu.    :36.00 3rd Qu.    :3.000
#> Max.       :999.0 Max.       :3.0000 Max.       :72.00 Max.       :4.000
#> new_car      used_car      furniture      radio.tv
#> Min.      :0.0000 Min.      :0.00000 Min.      :0.0000 Min.      :0.0000
#> 1st Qu.    :0.0000 1st Qu.    :0.00000 1st Qu.    :0.0000 1st Qu.    :0.0000
#> Median     :0.0000 Median     :0.00000 Median     :0.0000 Median     :0.0000

```

```

#> Mean :0.2967 Mean :0.05667 Mean :0.1933 Mean :0.2067
#> 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.0000
#> Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.0000
#> education retraining amount sav_acct
#> Min. :0.00000 Min. :0.0000 Min. : 433 Min. :0.0000
#> 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.: 1352 1st Qu.:0.0000
#> Median :0.00000 Median :0.0000 Median : 2574 Median :0.0000
#> Mean :0.07333 Mean :0.1133 Mean : 3938 Mean :0.6733
#> 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.: 5142 3rd Qu.:1.0000
#> Max. :1.00000 Max. :1.0000 Max. :18424 Max. :4.0000
#> employment install_rate male_div male_single
#> Min. :0.00 Min. :1.000 Min. :0.00000 Min. :0.0000
#> 1st Qu.:1.00 1st Qu.:2.000 1st Qu.:0.00000 1st Qu.:0.0000
#> Median :2.00 Median :4.000 Median :0.00000 Median :0.0000
#> Mean :2.17 Mean :3.097 Mean :0.06667 Mean :0.4867
#> 3rd Qu.:3.00 3rd Qu.:4.000 3rd Qu.:0.00000 3rd Qu.:1.0000
#> Max. :4.00 Max. :4.000 Max. :1.00000 Max. :1.0000
#> male_mar_or_wid co.applicant guarantor present_resident
#> Min. :0.00000 Min. :0.00 Min. :0.00000 Min. :1.00
#> 1st Qu.:0.00000 1st Qu.:0.00 1st Qu.:0.00000 1st Qu.:2.00
#> Median :0.00000 Median :0.00 Median :0.00000 Median :3.00
#> Mean :0.08333 Mean :0.06 Mean :0.03333 Mean :2.85
#> 3rd Qu.:0.00000 3rd Qu.:0.00 3rd Qu.:0.00000 3rd Qu.:4.00
#> Max. :1.00000 Max. :1.00 Max. :1.00000 Max. :4.00
#> real_estate prop_unkn_none age other_install
#> Min. :0.0 Min. :0.0000 Min. :19.00 Min. :0.0000
#> 1st Qu.:0.0 1st Qu.:0.0000 1st Qu.:25.00 1st Qu.:0.0000
#> Median :0.0 Median :0.0000 Median :31.00 Median :0.0000
#> Mean :0.2 Mean :0.2233 Mean :33.96 Mean :0.2533
#> 3rd Qu.:0.0 3rd Qu.:0.0000 3rd Qu.:40.00 3rd Qu.:1.0000
#> Max. :1.0 Max. :1.0000 Max. :74.00 Max. :1.0000
#> rent own_res num_credits job
#> Min. :0.0000 Min. :0.00 Min. :1.000 Min. :0.000
#> 1st Qu.:0.0000 1st Qu.:0.00 1st Qu.:1.000 1st Qu.:2.000
#> Median :0.0000 Median :1.00 Median :1.000 Median :2.000
#> Mean :0.2333 Mean :0.62 Mean :1.367 Mean :1.937
#> 3rd Qu.:0.0000 3rd Qu.:1.00 3rd Qu.:2.000 3rd Qu.:2.000
#> Max. :1.0000 Max. :1.00 Max. :4.000 Max. :3.000
#> num_dependents telephone foreign response
#> Min. :1.000 Min. :0.0000 Min. :0.00000 Min. :0
#> 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0
#> Median :1.000 Median :0.0000 Median :0.00000 Median :0
#> Mean :1.153 Mean :0.3767 Mean :0.01333 Mean :0
#> 3rd Qu.:1.000 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0
#> Max. :2.000 Max. :1.0000 Max. :1.00000 Max. :0

```

Transforming all categorical variables to factor:

```

categorical <- c("chk_acct", "new_car", "used_car", "furniture",
  "radio.tv", "education", "retraining", "sav_acct",
  "employment", "male_div", "male_single", "male_mar_or_wid",
  "co.applicant", "guarantor", "present_resident", "real_estate",
  "prop_unkn_none", "other_install", "rent", "own_res", "job",
  "telephone", "foreign")

```

```

credit$response <- as.factor(credit$response)
names(credit)
#> [1] "obs."          "chk_acct"      "duration"      "history"
#> [5] "new_car"       "used_car"      "furniture"     "radio.tv"
#> [9] "education"     "retraining"    "amount"        "sav_acct"
#> [13] "employment"    "install_rate"  "male_div"      "male_single"
#> [17] "male_mar_or_wid" "co.applicant"  "guarantor"     "present_resident"
#> [21] "real_estate"   "prop_unkn_none" "age"           "other_install"
#> [25] "rent"          "own_res"       "num_credits"   "job"
#> [29] "num_dependents" "telephone"     "foreign"       "response"

```

```

plt <- function(col) {
  print(col)
  var <- credit %>% dplyr::select(col)
  ggplot(credit, aes(y=var[,1], fill=response)) + geom_bar()
}

```

```

for (i in 1:length(categorical)) {
  plt(categorical[i])
}
#> [1] "chk_acct"
#> [1] "new_car"
#> [1] "used_car"
#> [1] "furniture"
#> [1] "radio.tv"
#> [1] "education"
#> [1] "retraining"
#> [1] "sav_acct"
#> [1] "employment"
#> [1] "male_div"
#> [1] "male_single"
#> [1] "male_mar_or_wid"
#> [1] "co.applicant"
#> [1] "guarantor"
#> [1] "present_resident"
#> [1] "real_estate"
#> [1] "prop_unkn_none"
#> [1] "other_install"
#> [1] "rent"
#> [1] "own_res"
#> [1] "job"
#> [1] "telephone"
#> [1] "foreign"

```