Homework_4

Anthony Munoz

4/26/2020

Contents

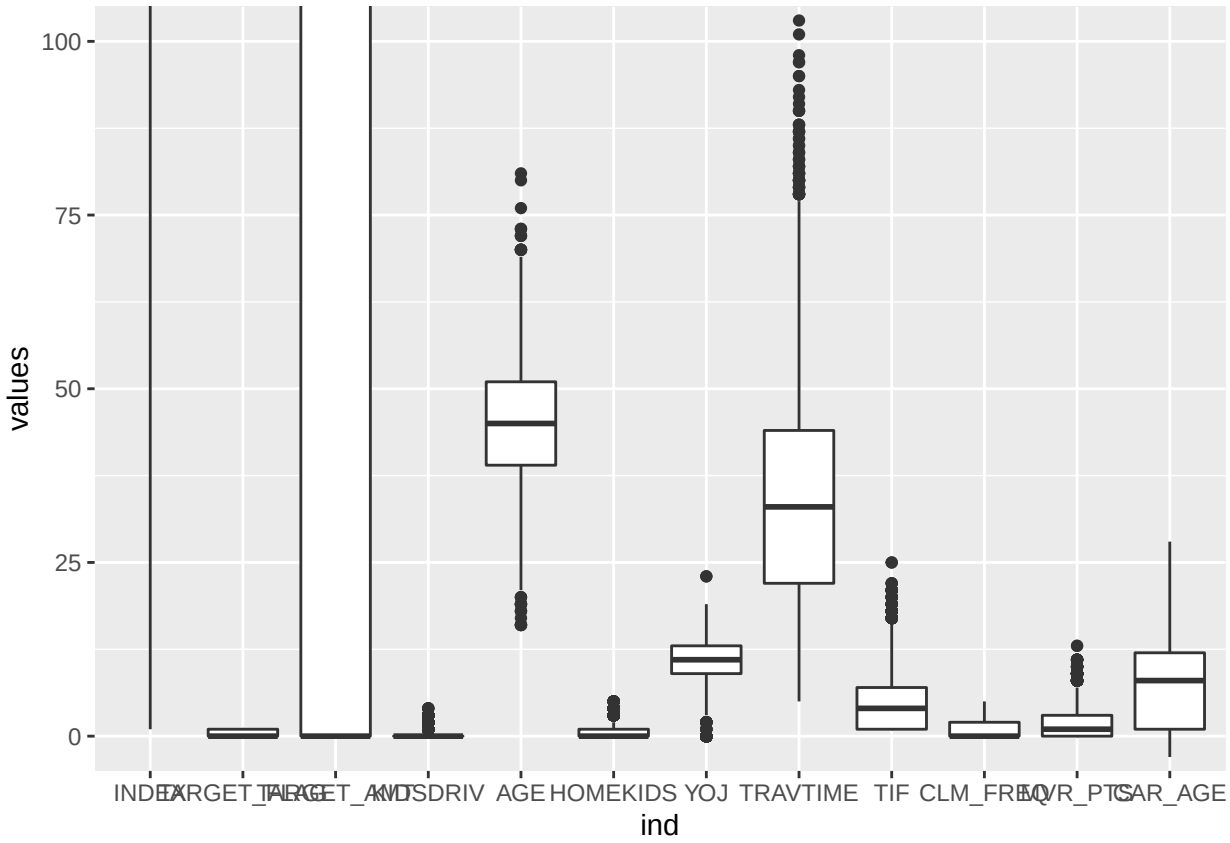
Data Exploration

```
## 'data.frame':    8161 obs. of  26 variables:
##  $ INDEX      : int  1 2 4 5 6 7 8 11 12 13 ...
##  $ TARGET_FLAG: int  0 0 0 0 0 1 0 1 1 0 ...
##  $ TARGET_AMT : num  0 0 0 0 0 ...
##  $ KIDSDRIV   : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ AGE        : int  60 43 35 51 50 34 54 37 34 50 ...
##  $ HOMEKIDS   : int  0 0 1 0 0 1 0 2 0 0 ...
##  $ YOJ        : int  11 11 10 14 NA 12 NA NA 10 7 ...
##  $ INCOME     : Factor w/ 6613 levels "","$0","$1,007",..: 5033 6292 1250 1 509 746 1488 315 4765 28
##  $ PARENT1    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...
##  $ HOME_VAL   : Factor w/ 5107 levels "","$0","$100,093",..: 2 3259 348 3917 3034 2 1 4167 2 2 ...
##  $ MSTATUS    : Factor w/ 2 levels "Yes","z_No": 2 2 1 1 1 2 1 1 2 2 ...
##  $ SEX        : Factor w/ 2 levels "M","z_F": 1 1 2 1 2 2 2 1 2 1 ...
##  $ EDUCATION  : Factor w/ 5 levels "<High School",..: 4 5 5 1 4 2 1 2 2 2 ...
##  $ JOB        : Factor w/ 9 levels "","Clerical",..: 7 9 2 9 3 9 9 9 2 7 ...
##  $ TRAVTIME   : int  14 22 5 32 36 46 33 44 34 48 ...
##  $ CAR_USE    : Factor w/ 2 levels "Commercial","Private": 2 1 2 2 2 1 2 1 2 1 ...
##  $ BLUEBOOK   : Factor w/ 2789 levels "$1,500","$1,520",..: 434 503 2212 553 802 746 2672 701 135 85
##  $ TIF        : int  11 1 4 7 1 1 1 1 1 7 ...
##  $ CAR_TYPE   : Factor w/ 6 levels "Minivan","Panel Truck",..: 1 1 6 1 6 4 6 5 6 5 ...
##  $ RED_CAR    : Factor w/ 2 levels "no","yes": 2 2 1 2 1 1 1 2 1 1 ...
##  $ OLDCLAIM   : Factor w/ 2857 levels "$0","$1,000",..: 1449 1 1311 1 432 1 1 510 1 1 ...
##  $ CLM_FREQ   : int  2 0 2 0 2 0 0 1 0 0 ...
##  $ REVOKED    : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 1 2 1 1 ...
##  $ MVR_PTS    : int  3 0 3 0 3 0 0 10 0 1 ...
##  $ CAR_AGE    : int  18 1 10 6 17 7 1 7 1 17 ...
##  $ URBANICITY : Factor w/ 2 levels "Highly Urban/ Urban",..: 1 1 1 1 1 1 1 1 1 2 ...
```

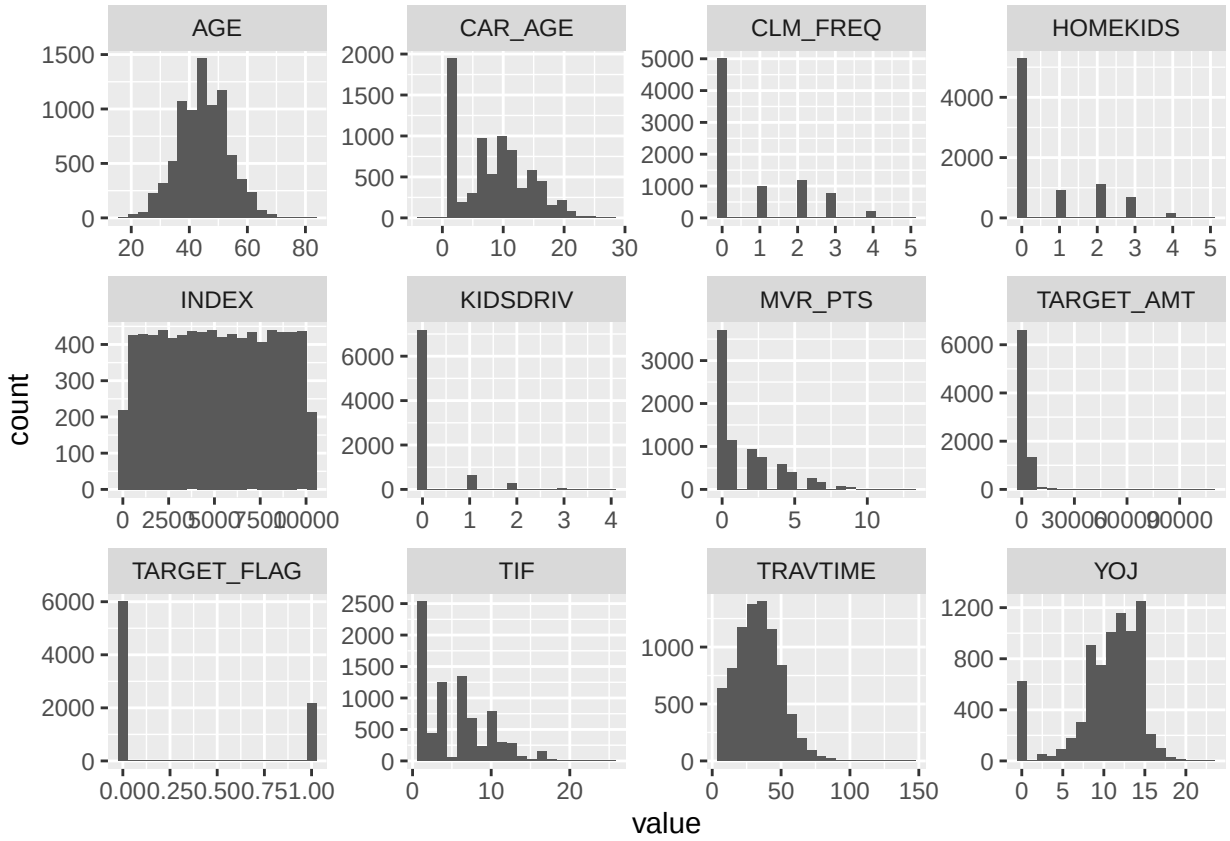
```
##      INDEX        TARGET_FLAG       TARGET_AMT        KIDSDRIV
##  Min.   :    1   Min.   :0.0000   Min.   :     0   Min.   :0.0000
##  1st Qu.: 2559   1st Qu.:0.0000   1st Qu.:     0   1st Qu.:0.0000
##  Median : 5133   Median :0.0000   Median :     0   Median :0.0000
##  Mean   : 5152   Mean   :0.2638   Mean   :  1504   Mean   :0.1711
##  3rd Qu.: 7745   3rd Qu.:1.0000   3rd Qu.:  1036   3rd Qu.:0.0000
##  Max.   :10302   Max.   :1.0000   Max.   :107586   Max.   :4.0000
##
##       AGE           HOMEKIDS           YOJ            INCOME     PARENT1
##  Min.   :16.00   Min.   :0.0000   Min.   : 0.0   $0      : 615   No :7084
##  1st Qu.:39.00   1st Qu.:0.0000   1st Qu.: 9.0           : 445   Yes:1077
##  Median :45.00   Median :0.0000   Median :11.0   $26,840 :   4
##  Mean   :44.79   Mean   :0.7212   Mean   :10.5   $48,509 :   4
##  3rd Qu.:51.00   3rd Qu.:1.0000   3rd Qu.:13.0   $61,790 :   4
##  Max.   :81.00   Max.   :5.0000   Max.   :23.0   $107,375:   3
##  NA's   :6                        NA's   :454    (Other) :7086
##      HOME_VAL    MSTATUS      SEX               EDUCATION
##  $0      :2294   Yes :4894   M  :3786   <High School :1203
##          : 464   z_No:3267   z_F:4375   Bachelors    :2242
##  $111,129:   3                          Masters      :1658
##  $115,249:   3                          PhD          : 728
##  $123,109:   3                          z_High School:2330
##  $153,061:   3
##  (Other) :5391
##             JOB          TRAVTIME            CAR_USE        BLUEBOOK
##  z_Blue Collar:1825   Min.   :  5.00   Commercial:3029   $1,500 : 157
##  Clerical     :1271   1st Qu.: 22.00   Private   :5132   $6,000 :  34
##  Professional :1117   Median : 33.00                     $5,800 :  33
##  Manager      : 988   Mean   : 33.49                     $6,200 :  33
##  Lawyer       : 835   3rd Qu.: 44.00                     $6,400 :  31
##  Student      : 712   Max.   :142.00                     $5,900 :  30
##  (Other)      :1413                                      (Other):7843
##       TIF               CAR_TYPE    RED_CAR       OLDCLAIM       CLM_FREQ
##  Min.   : 1.000   Minivan    :2145   no :5783   $0     :5009   Min.   :0.0000
##  1st Qu.: 1.000   Panel Truck: 676   yes:2378   $1,310 :   4   1st Qu.:0.0000
##  Median : 4.000   Pickup     :1389              $1,391 :   4   Median :0.0000
##  Mean   : 5.351   Sports Car : 907              $4,263 :   4   Mean   :0.7986
##  3rd Qu.: 7.000   Van        : 750              $1,105 :   3   3rd Qu.:2.0000
##  Max.   :25.000   z_SUV      :2294              $1,332 :   3   Max.   :5.0000
##                                                 (Other):3134
##  REVOKED       MVR_PTS          CAR_AGE                      URBANICITY
##  No :7161   Min.   : 0.000   Min.   :-3.000   Highly Urban/ Urban  :6492
##  Yes:1000   1st Qu.: 0.000   1st Qu.: 1.000   z_Highly Rural/ Rural:1669
##             Median : 1.000   Median : 8.000
##             Mean   : 1.696   Mean   : 8.328
##             3rd Qu.: 3.000   3rd Qu.:12.000
##             Max.   :13.000   Max.   :28.000
##                              NA's   :510
```

```
## Warning in stack.data.frame(data): non-vector columns will be ignored
```

```
## Warning: Removed 970 rows containing non-finite values (stat_boxplot).
```




```
## Warning: Removed 970 rows containing non-finite values (stat_bin).
```

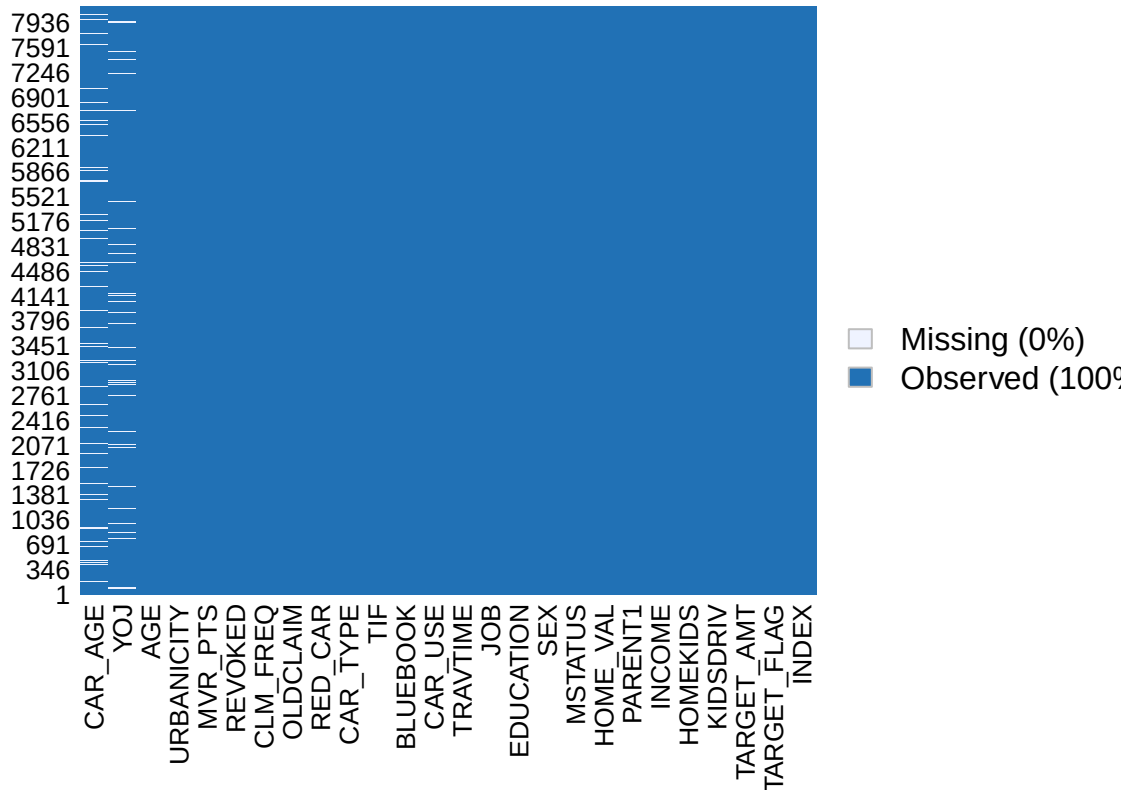


From the graph, we can see that just the Age variable is normally distributed. TRAVELTIME,YOJ, CAR_AGE seem distributed but they are skewed data, which will require some data transformation to fix them.

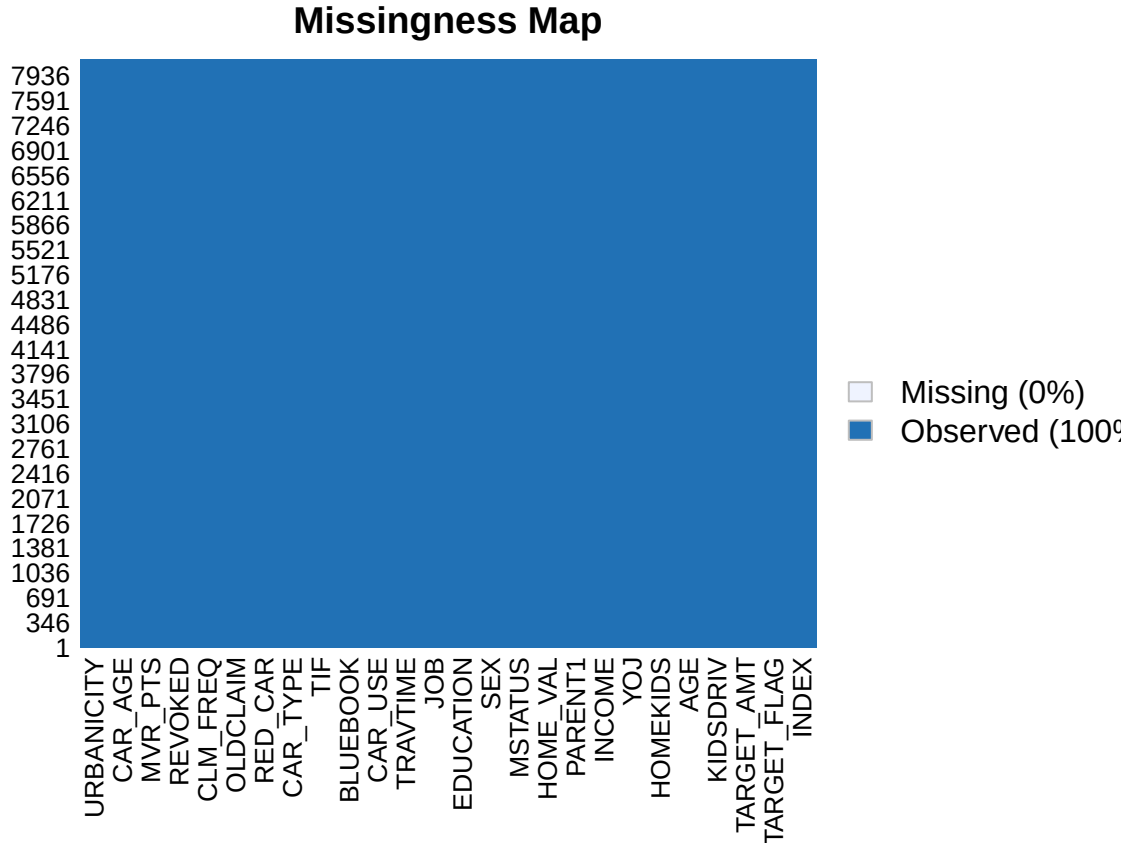
Data Preparation

```
##       INDEX TARGET_FLAG  TARGET_AMT    KIDSDRIV         AGE    HOMEKIDS
##   "integer"   "integer"   "numeric"   "integer"   "integer"   "integer"
##         YOJ      INCOME     PARENT1    HOME_VAL     MSTATUS         SEX
##   "integer"    "factor"    "factor"    "factor"    "factor"    "factor"
##   EDUCATION         JOB    TRAVTIME     CAR_USE    BLUEBOOK         TIF
##    "factor"    "factor"   "integer"    "factor"    "factor"   "integer"
##    CAR_TYPE     RED_CAR    OLDCLAIM    CLM_FREQ     REVOKED     MVR_PTS
##    "factor"    "factor"    "factor"   "integer"    "factor"   "integer"
##     CAR_AGE  URBANICITY
##   "integer"    "factor"
```

```
##  [1] "INDEX"       "TARGET_FLAG" "TARGET_AMT"  "KIDSDRIV"    "AGE"
##  [6] "HOMEKIDS"    "YOJ"         "INCOME"      "PARENT1"     "HOME_VAL"
## [11] "MSTATUS"     "SEX"         "EDUCATION"   "JOB"         "TRAVTIME"
## [16] "CAR_USE"     "BLUEBOOK"    "TIF"         "CAR_TYPE"    "RED_CAR"
## [21] "OLDCLAIM"    "CLM_FREQ"    "REVOKED"     "MVR_PTS"     "CAR_AGE"
## [26] "URBANICITY"
```




```
##       INDEX TARGET_FLAG  TARGET_AMT    KIDSDRIV         AGE    HOMEKIDS
##           0           0           0           0           6           0
##         YOJ      INCOME     PARENT1    HOME_VAL     MSTATUS         SEX
##         454           0           0           0           0           0
##   EDUCATION         JOB    TRAVTIME     CAR_USE    BLUEBOOK         TIF
##           0           0           0           0           0           0
##    CAR_TYPE     RED_CAR    OLDCLAIM    CLM_FREQ     REVOKED     MVR_PTS
##           0           0           0           0           0           0
##     CAR_AGE  URBANICITY
##         510           0
```



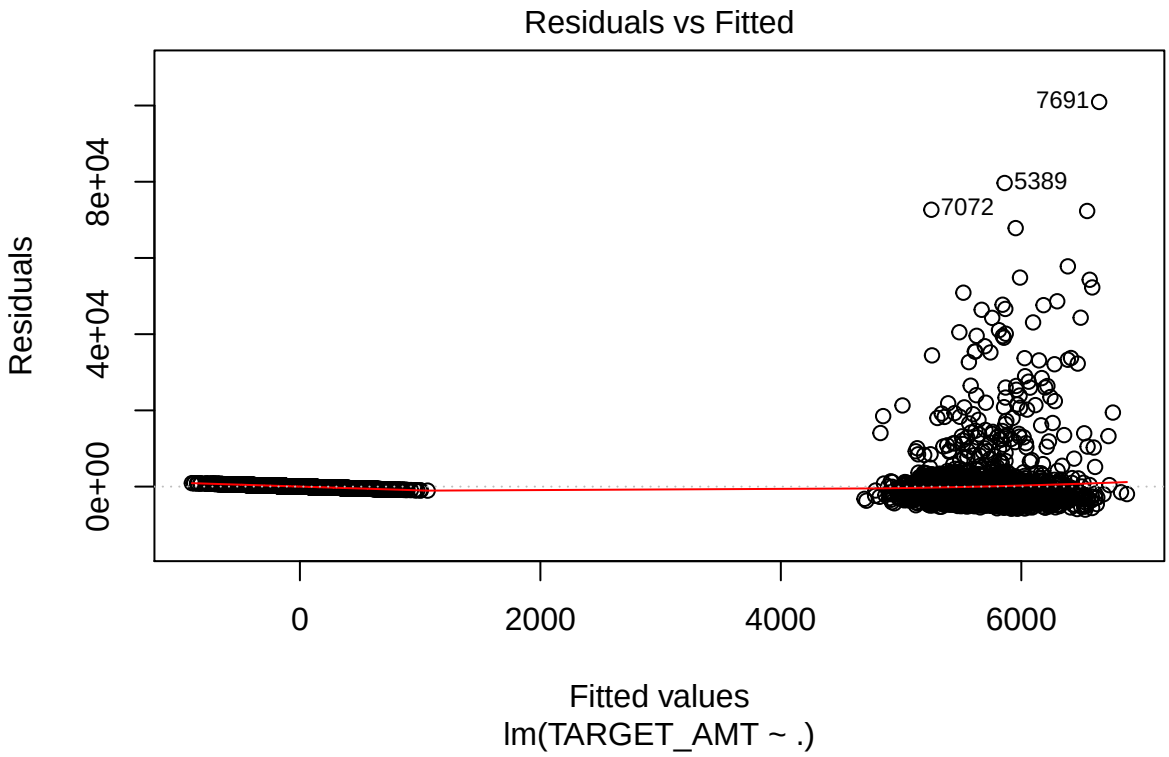
We can see some variables are missing some observation. in order to work with them, we decide to impute the data by the mean of those variables.

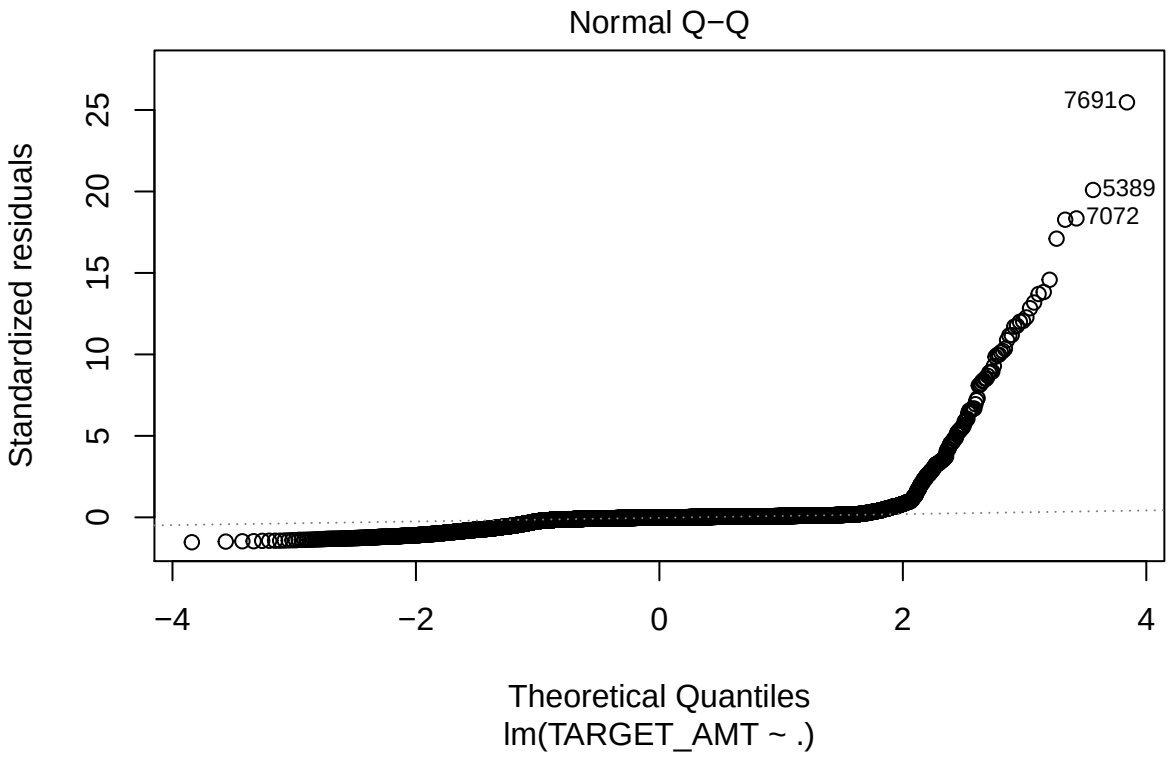
Model

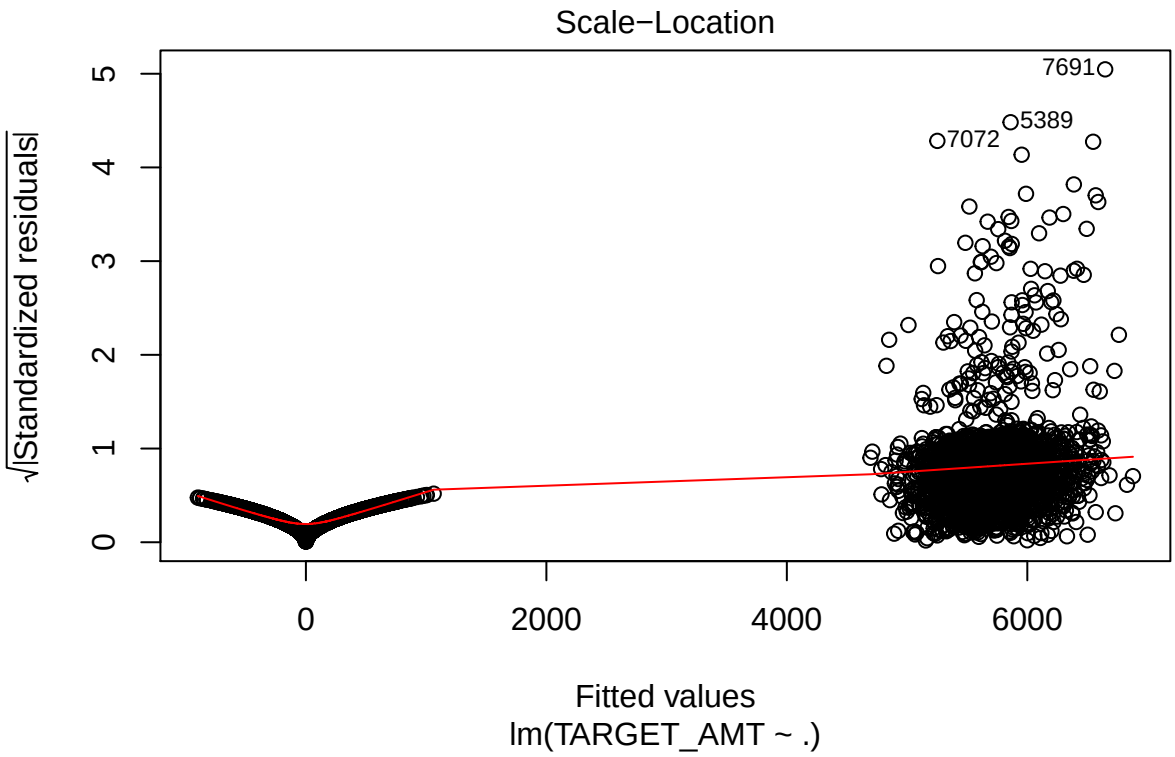
Multiple Linear Regression Model

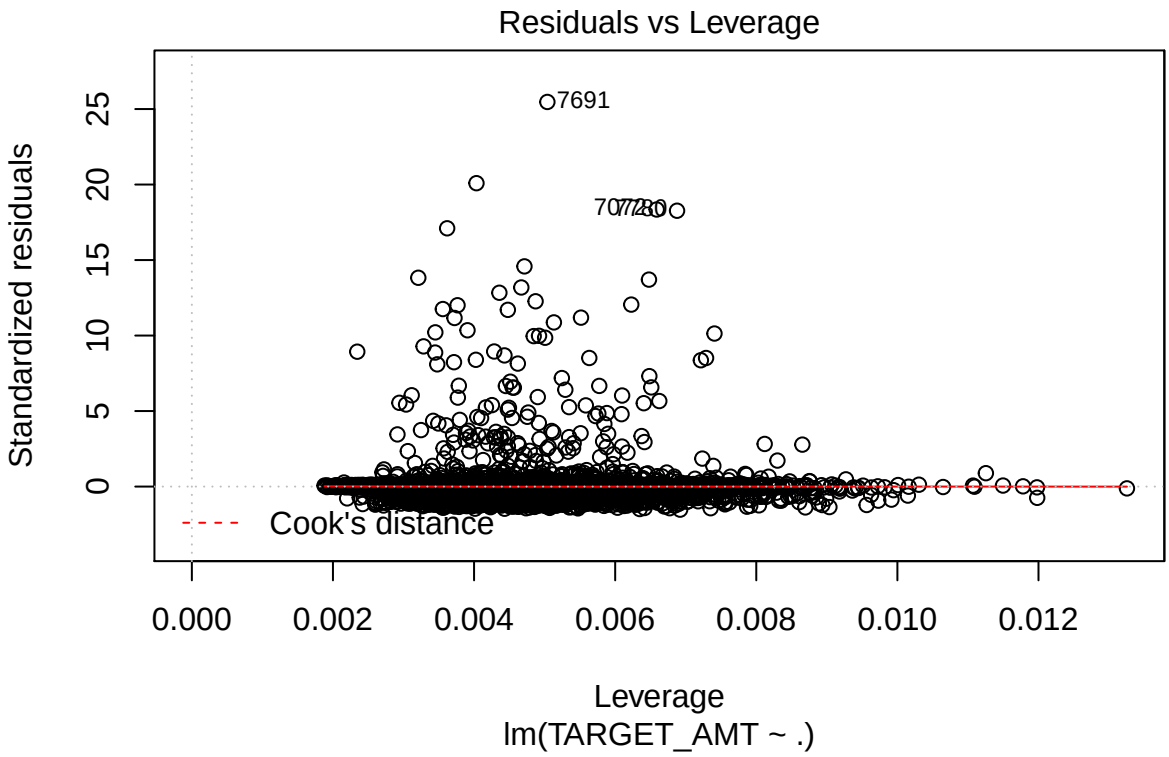
```
##
## Call:
## lm(formula = TARGET_AMT ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -6048   -407    -38    194 100939
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -9.958e+01  4.723e+02  -0.211  0.83301
## TARGET_FLAG          5.678e+03  1.132e+02  50.175  < 2e-16 ***
## KIDSDRIV            -3.299e+01  9.916e+01  -0.333  0.73935
## AGE                  1.013e+01  6.102e+00   1.660  0.09696 .
## HOMEKIDS             4.142e+01  5.715e+01   0.725  0.46867
## YOJ                  8.263e+00  1.339e+01   0.617  0.53706
## INCOME               9.127e-04  2.347e-02   0.039  0.96898
## PARENT1              1.634e+02  1.765e+02   0.926  0.35463
```

```
## HOME_VAL             1.008e-03  3.054e-02   0.033  0.97368
## MSTATUS             -1.290e+02  1.175e+02  -1.098  0.27241
## SEX                  6.092e+01  1.497e+02   0.407  0.68407
## EDUCATION           -8.411e+01  1.462e+02  -0.575  0.56517
## JOBClerical         -2.315e+02  2.485e+02  -0.932  0.35143
## JOBDoctor           -1.255e+02  3.302e+02  -0.380  0.70382
## JOBHome Maker       -2.208e+02  2.800e+02  -0.789  0.43038
## JOBLawyer            5.862e+01  2.509e+02   0.234  0.81529
## JOBManager          -2.123e+02  2.340e+02  -0.908  0.36416
## JOBProfessional      3.959e+01  2.329e+02   0.170  0.86502
## JOBStudent          -3.869e+02  2.767e+02  -1.399  0.16197
## JOBz_Blue Collar    -1.734e+02  2.318e+02  -0.748  0.45436
## TRAVTIME             5.400e-01  2.824e+00   0.191  0.84838
## CAR_USE             -1.187e+02  1.443e+02  -0.822  0.41098
## BLUEBOOK            -1.680e-02  5.214e-02  -0.322  0.74726
## TIF                 -3.439e+00  1.068e+01  -0.322  0.74742
## CAR_TYPEPanel Truck  3.691e+02  2.219e+02   1.663  0.09636 .
## CAR_TYPEPickup      -8.904e+01  1.516e+02  -0.587  0.55690
## CAR_TYPESports Car  -3.085e+01  1.808e+02  -0.171  0.86457
## CAR_TYPEVan          2.707e+02  1.815e+02   1.491  0.13589
## CAR_TYPEz_SUV       -6.785e+01  1.454e+02  -0.467  0.64069
## RED_CAR             -4.467e+01  1.302e+02  -0.343  0.73161
## OLDCLAIM            -9.641e-02  7.313e-02  -1.318  0.18738
## CLM_FREQ             8.614e+00  5.513e+01   0.156  0.87584
## REVOKED             -3.120e+02  1.369e+02  -2.280  0.02265 *
## MVR_PTS              5.919e+01  2.297e+01   2.577  0.00998 **
## CAR_AGE             -1.649e+01  9.925e+00  -1.661  0.09673 .
## URBANICITY          -8.136e+00  1.267e+02  -0.064  0.94879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3973 on 8125 degrees of freedom
## Multiple R-squared:  0.2897, Adjusted R-squared:  0.2866
## F-statistic: 94.66 on 35 and 8125 DF,  p-value: < 2.2e-16
```

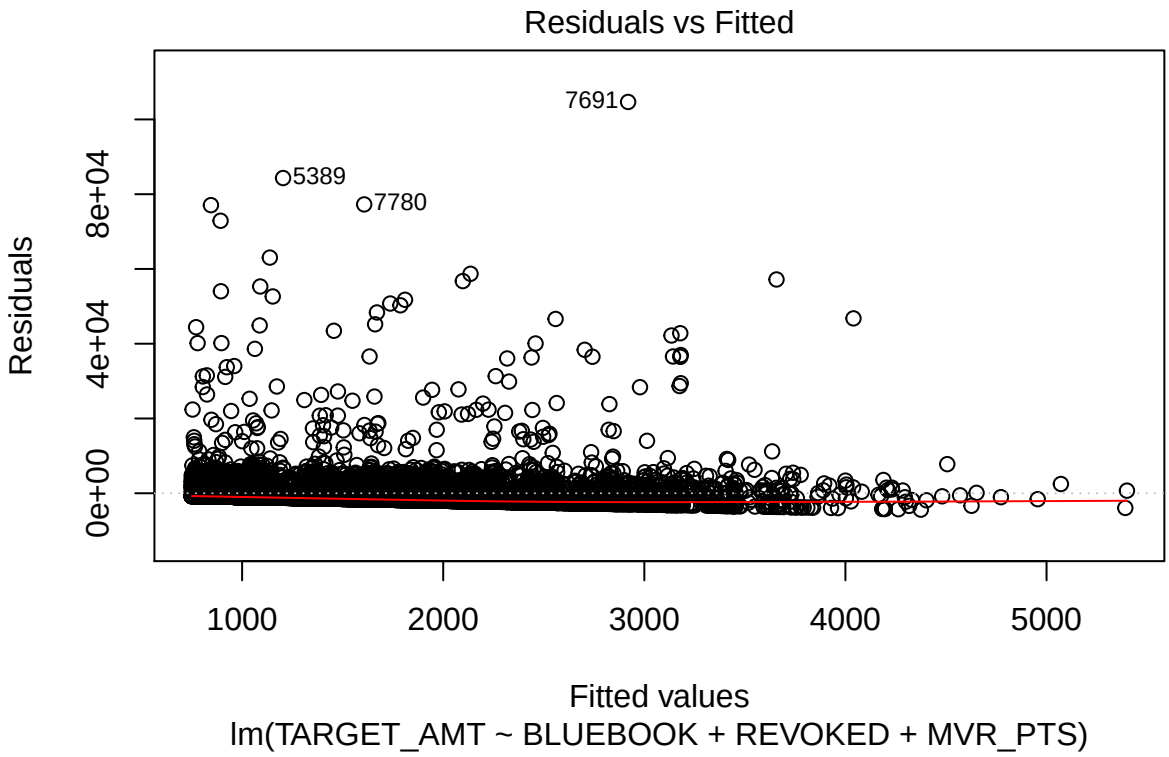

Residuals vs Fitted
Residuals
0e+00
4e+04
8e+04
7691
5389
7072
0
2000
4000
6000
Fitted values
lm(TARGET_AMT ~ .)

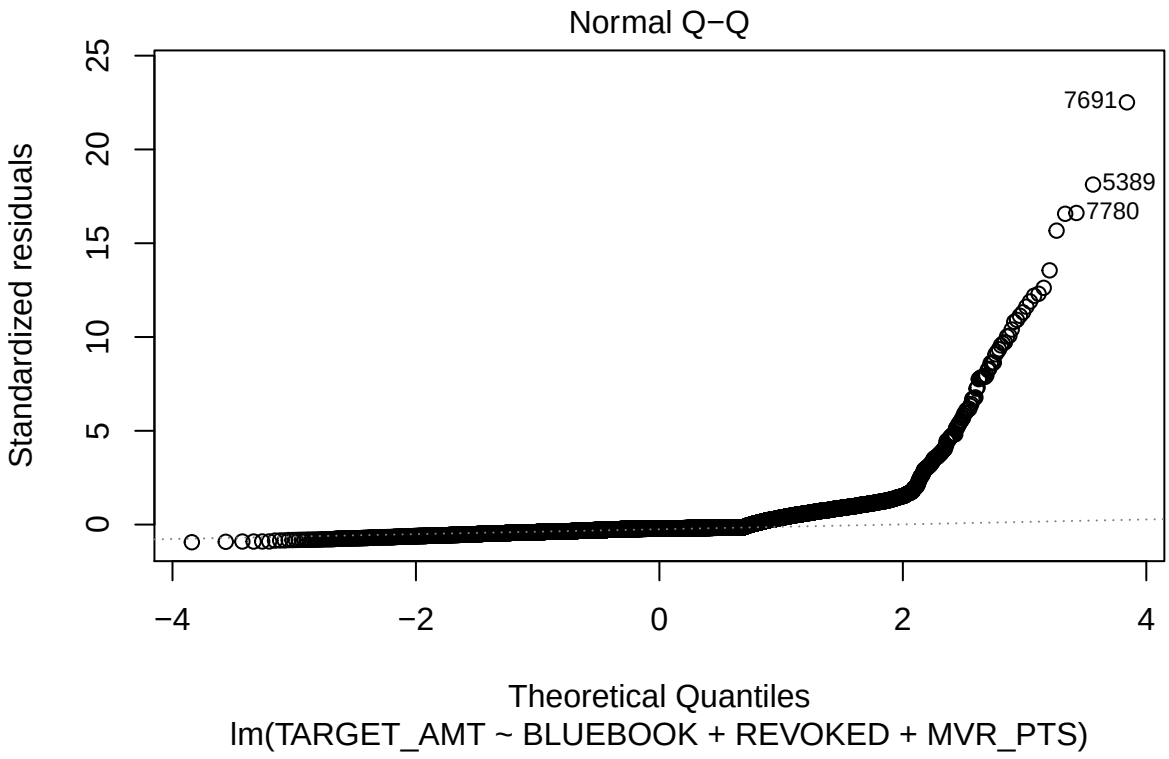

Normal Q−Q
Standardized residuals
0
5
10
15
20
25
7691
5389
7072
−4
−2
0
2
4
Theoretical Quantiles
lm(TARGET_AMT ~ .)

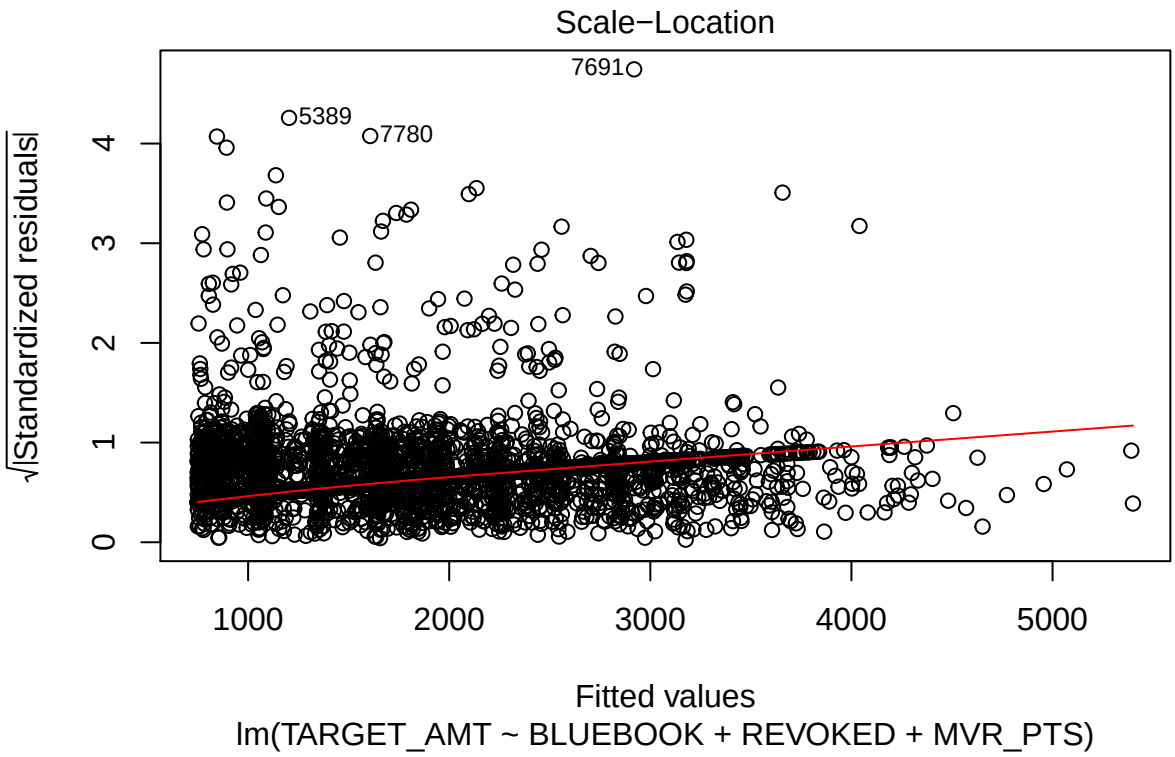

Scale−Location
√|Standardized residuals|
0
1
2
3
4
5
7691
5389
7072
0
2000
4000
6000
Fitted values
lm(TARGET_AMT ~ .)

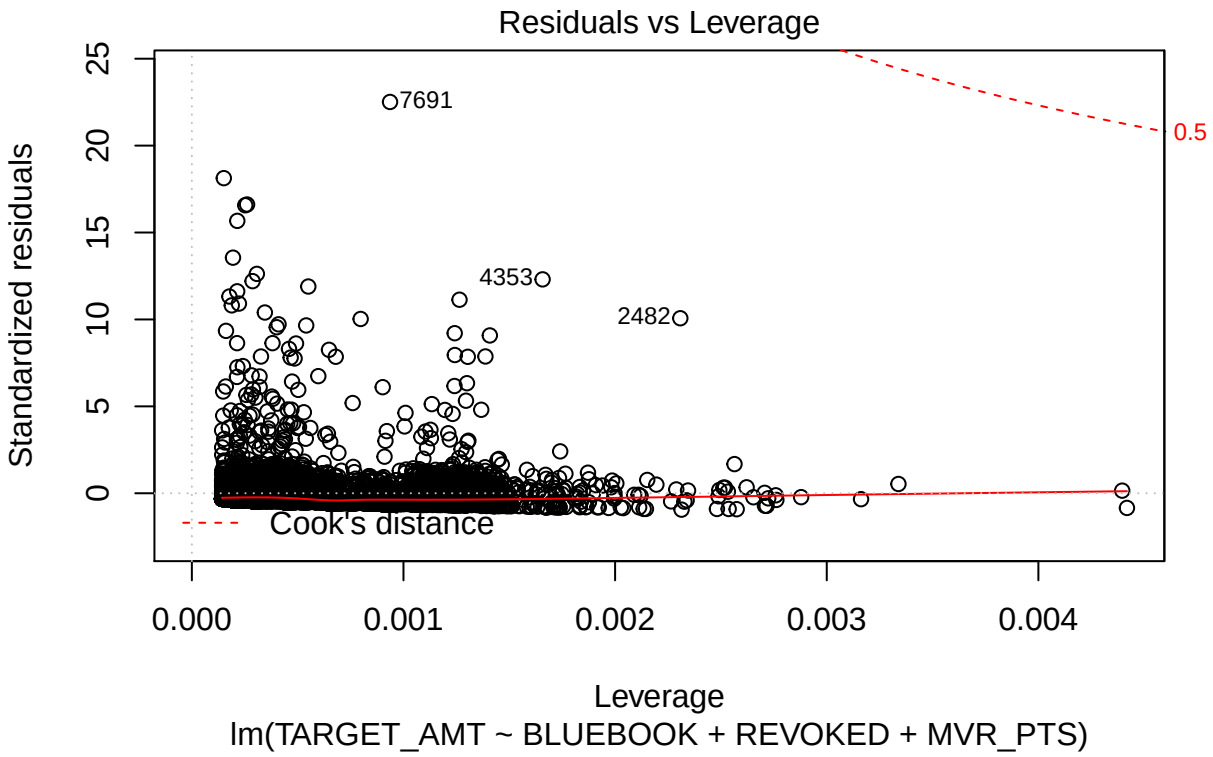



```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + REVOKED + MVR_PTS, data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -4375  -1566  -1001   -750 104667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 749.87577  100.30054   7.476 8.44e-14 ***
## BLUEBOOK      0.12301    0.05763   2.134   0.0328 *
## REVOKED     777.46752  157.25755   4.944 7.81e-07 ***
## MVR_PTS     295.65390   24.01748  12.310  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4652 on 8157 degrees of freedom
## Multiple R-squared:  0.02248,    Adjusted R-squared:  0.02212
## F-statistic: 62.54 on 3 and 8157 DF,  p-value: < 2.2e-16
```

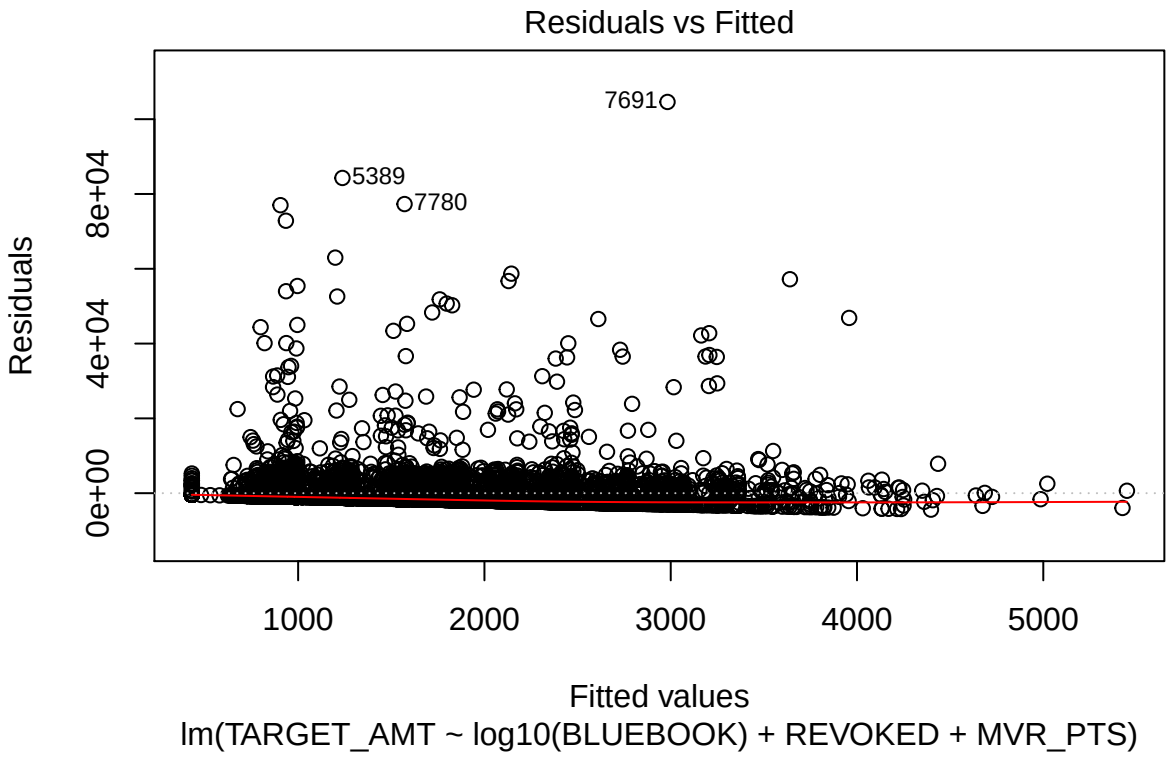

Residuals vs Fitted
7691
5389
7780
Residuals
0e+00
4e+04
8e+04
1000
2000
3000
4000
5000
Fitted values
lm(TARGET_AMT ~ BLUEBOOK + REVOKED + MVR_PTS)

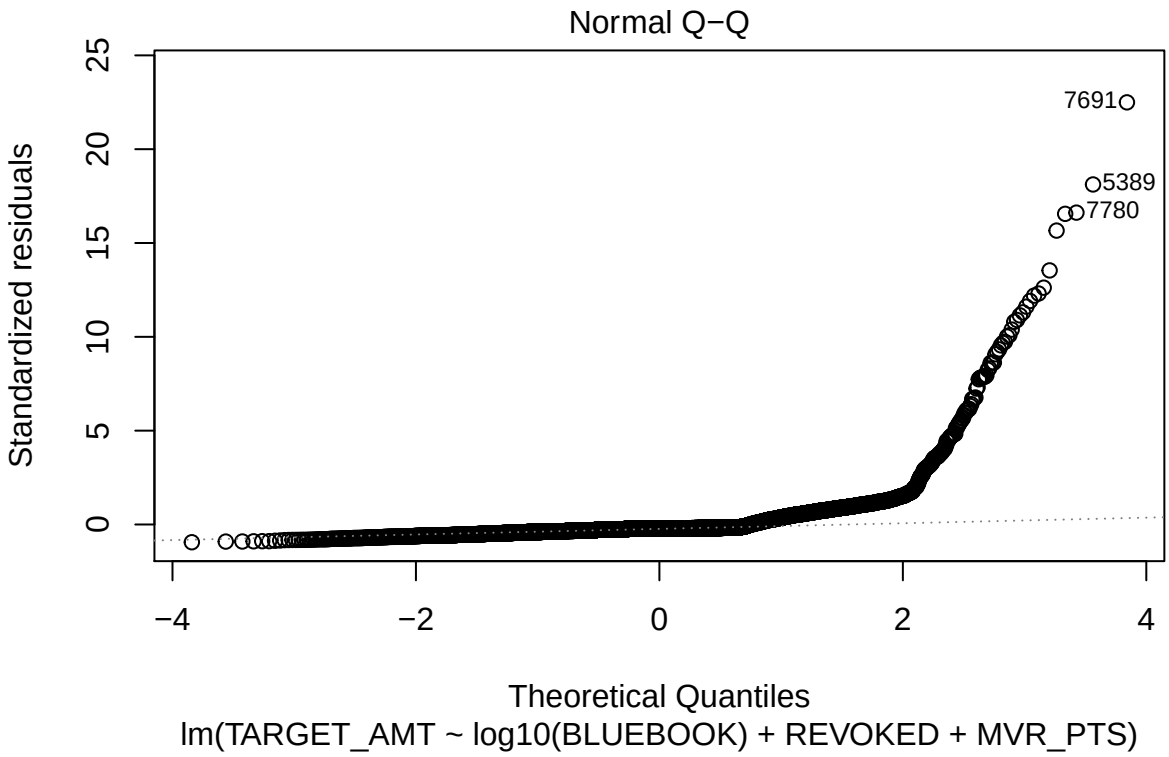

Normal Q−Q
Standardized residuals
0
5
10
15
20
25
7691
5389
7780
−4
−2
0
2
4
Theoretical Quantiles
lm(TARGET_AMT ~ BLUEBOOK + REVOKED + MVR_PTS)

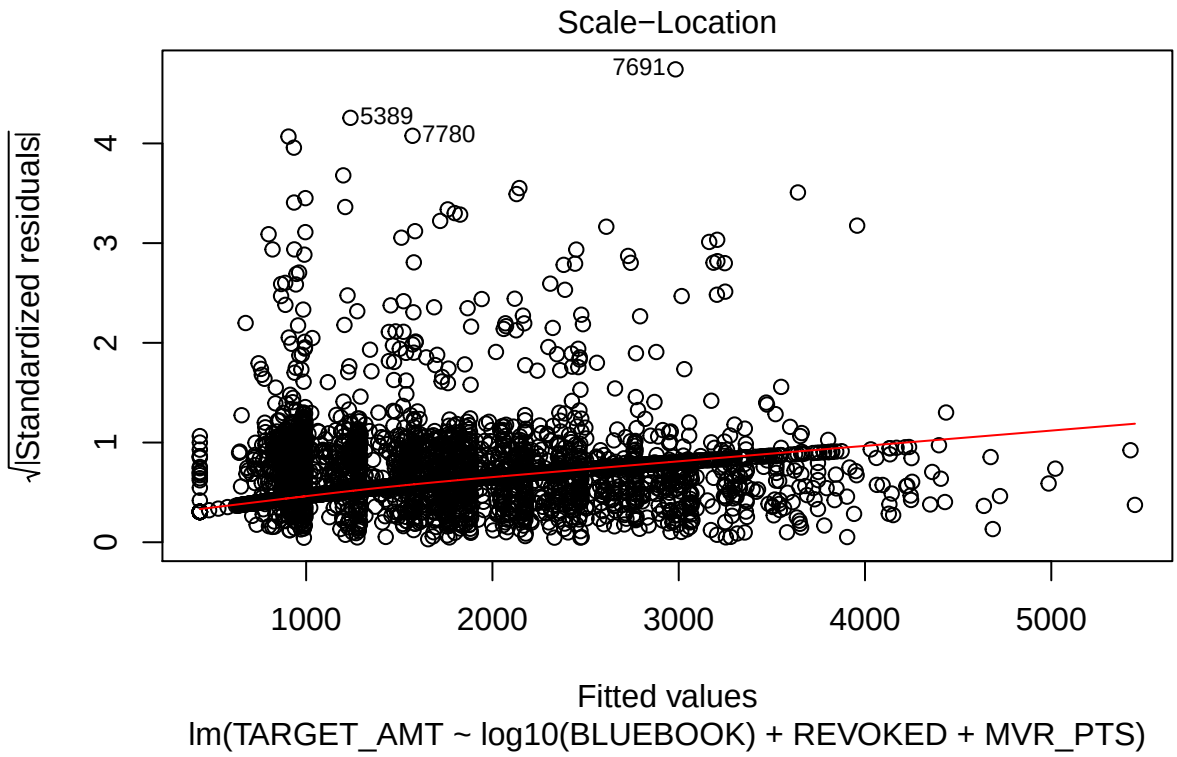

Scale−Location
7691
5389
7780
√|Standardized residuals|
0
1
2
3
4
1000
2000
3000
4000
5000
Fitted values
lm(TARGET_AMT ~ BLUEBOOK + REVOKED + MVR_PTS)

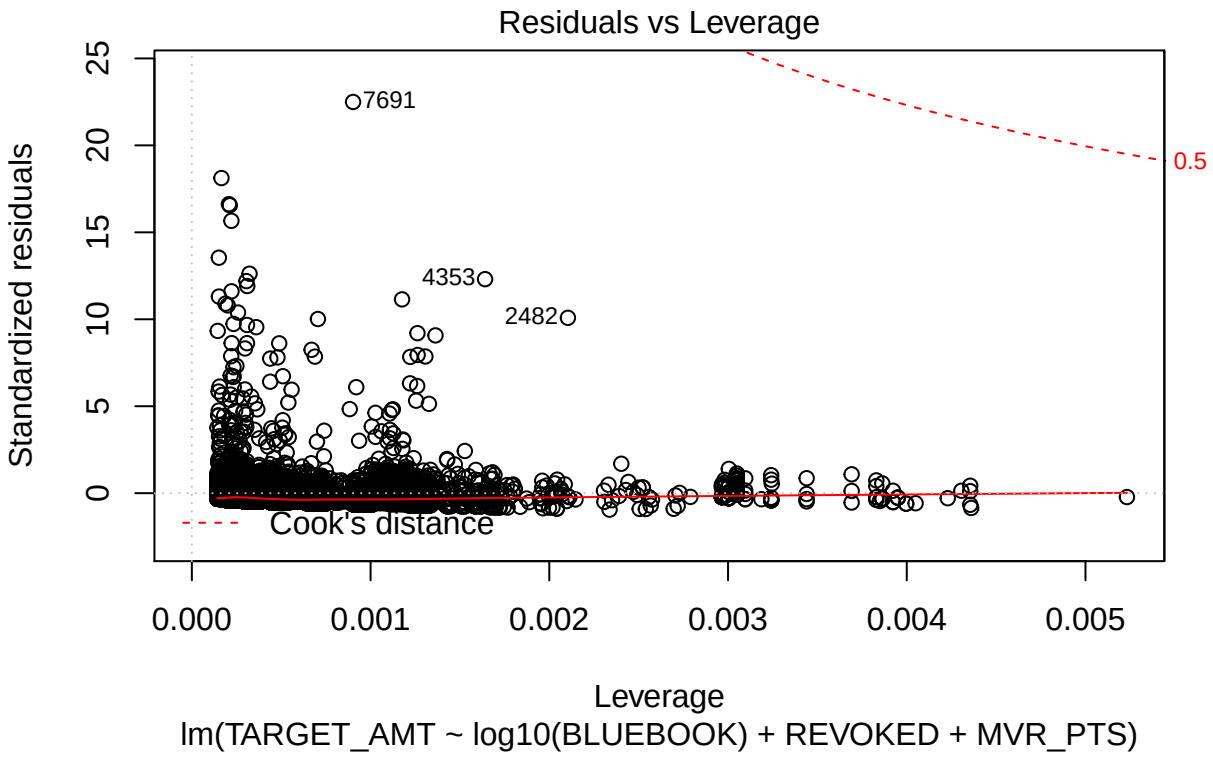



```
##
## Call:
## lm(formula = TARGET_AMT ~ log10(BLUEBOOK) + REVOKED + MVR_PTS,
##     data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -4397  -1559   -973   -620 104603
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       429.36     256.87   1.672   0.0947 .
## log10(BLUEBOOK)   164.72      85.33   1.930   0.0536 .
## REVOKED           777.04     157.27   4.941 7.93e-07 ***
## MVR_PTS           296.29      24.02  12.335  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4652 on 8157 degrees of freedom
## Multiple R-squared:  0.02238,    Adjusted R-squared:  0.02202
## F-statistic: 62.25 on 3 and 8157 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
7691
5389
7780
Residuals
0e+00
4e+04
8e+04
1000
2000
3000
4000
5000
Fitted values
lm(TARGET_AMT ~ log10(BLUEBOOK) + REVOKED + MVR_PTS)


Normal Q−Q
Standardized residuals
0
5
10
15
20
25
7691
5389
7780
−4
−2
0
2
4
Theoretical Quantiles
lm(TARGET_AMT ~ log10(BLUEBOOK) + REVOKED + MVR_PTS)

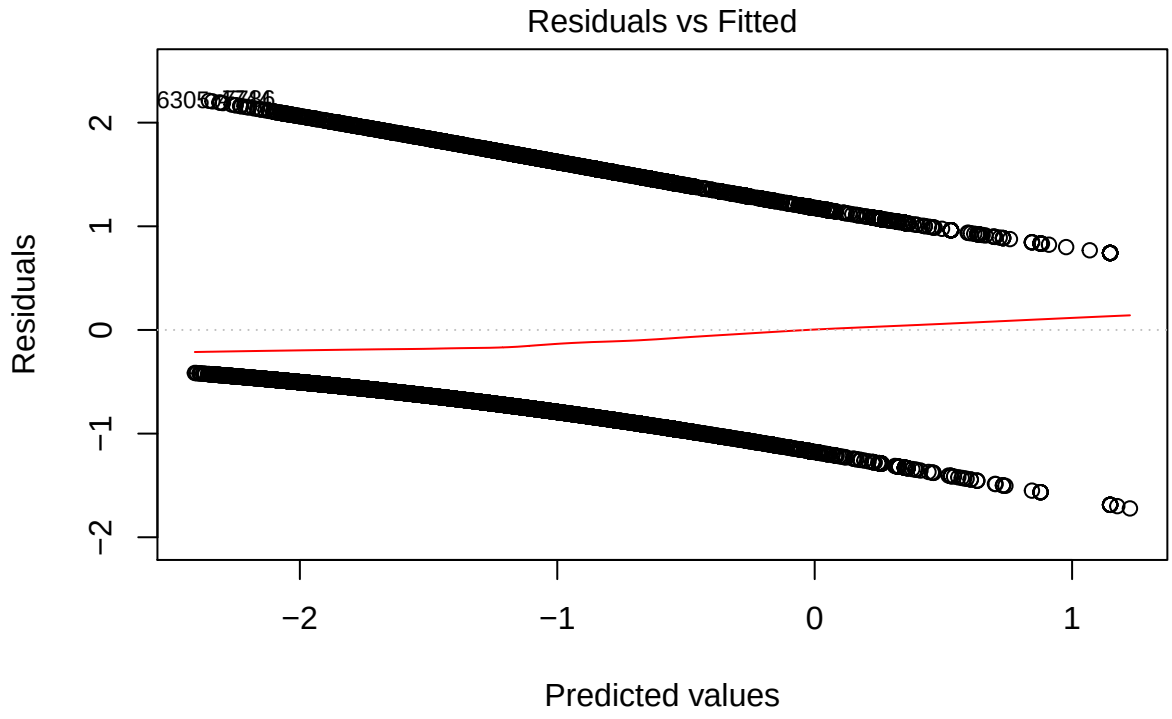

Scale−Location
7691
5389
7780
√|Standardized residuals|
0
1
2
3
4
1000
2000
3000
4000
5000
Fitted values
lm(TARGET_AMT ~ log10(BLUEBOOK) + REVOKED + MVR_PTS)



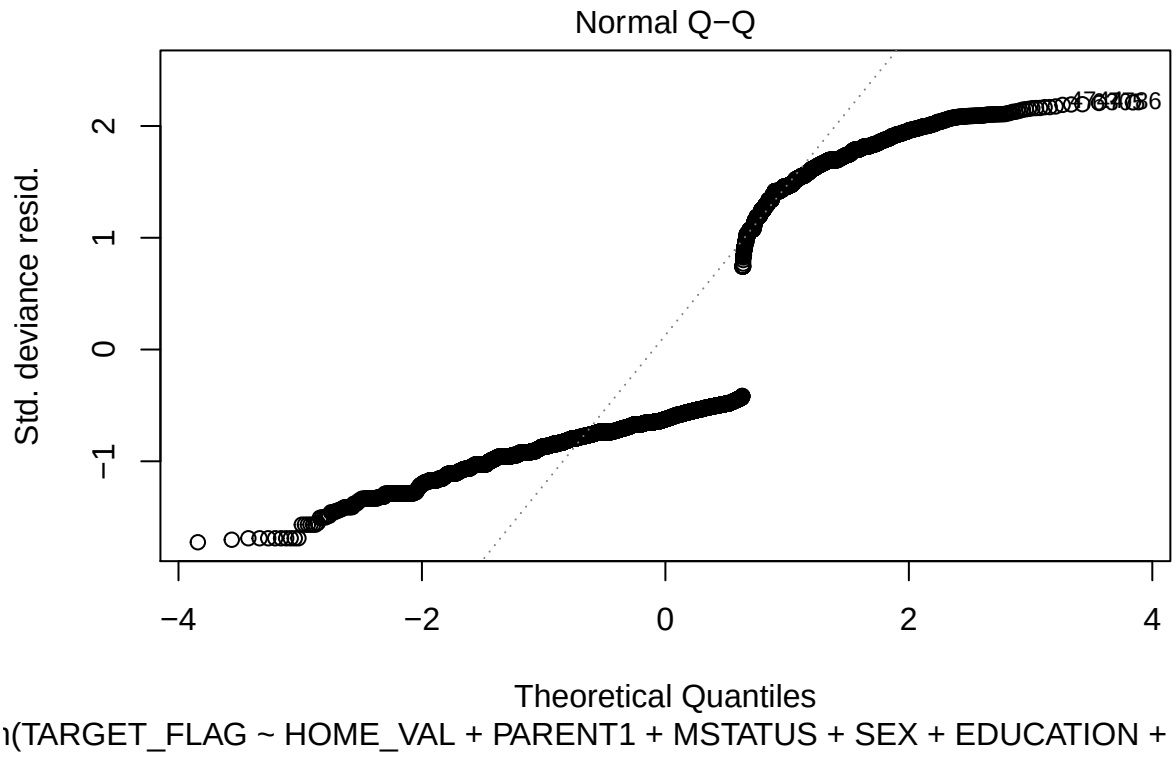
Logistic Regression Model

```
##
## Call:
## glm(formula = TARGET_FLAG ~ HOME_VAL + PARENT1 + MSTATUS + SEX +
##     EDUCATION + CAR_USE + RED_CAR + REVOKED, family = "binomial",
##     data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7219  -0.7749  -0.6236   1.0325   2.2111
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.312e-02  9.737e-02   0.135  0.89283
## HOME_VAL    -1.431e-04  1.772e-05  -8.073 6.87e-16 ***
## PARENT1      6.184e-01  8.061e-02   7.671 1.70e-14 ***
## MSTATUS     -2.097e-01  6.435e-02  -3.259  0.00112 **
## SEX         -3.037e-01  7.361e-02  -4.126 3.69e-05 ***
## EDUCATION   -3.808e-01  7.263e-02  -5.243 1.58e-07 ***
## CAR_USE     -7.968e-01  5.650e-02 -14.102  < 2e-16 ***
## RED_CAR      3.351e-02  7.847e-02   0.427  0.66934
## REVOKED      8.973e-01  7.238e-02  12.398  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

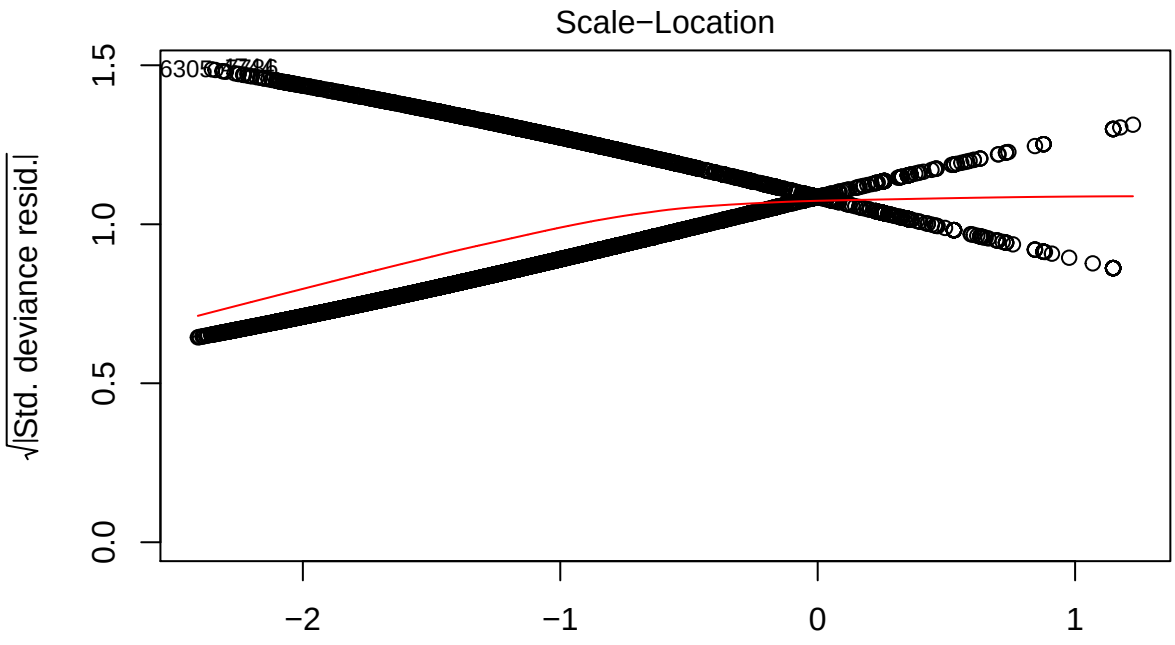
```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9418.0  on 8160  degrees of freedom
## Residual deviance: 8739.6  on 8152  degrees of freedom
## AIC: 8757.6
##
## Number of Fisher Scoring iterations: 4
```



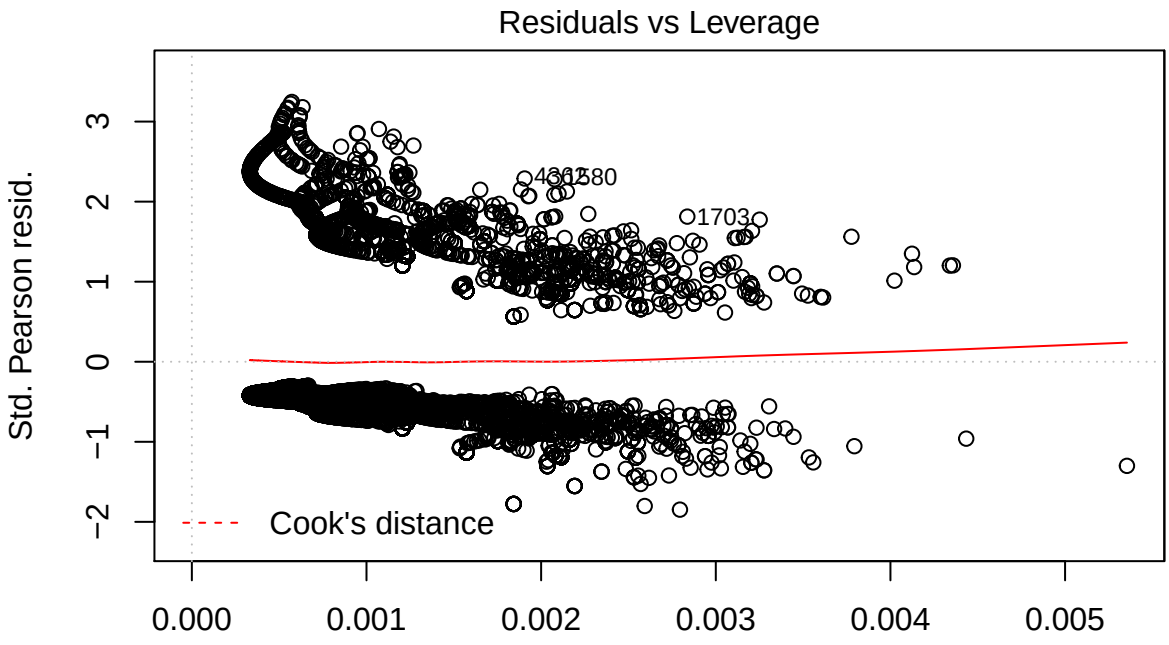
ı(TARGET_FLAG ~ HOME_VAL + PARENT1 + MSTATUS + SEX + EDUCATION + CAR_



ı(TARGET_FLAG ~ HOME_VAL + PARENT1 + MSTATUS + SEX + EDUCATION + CAR_



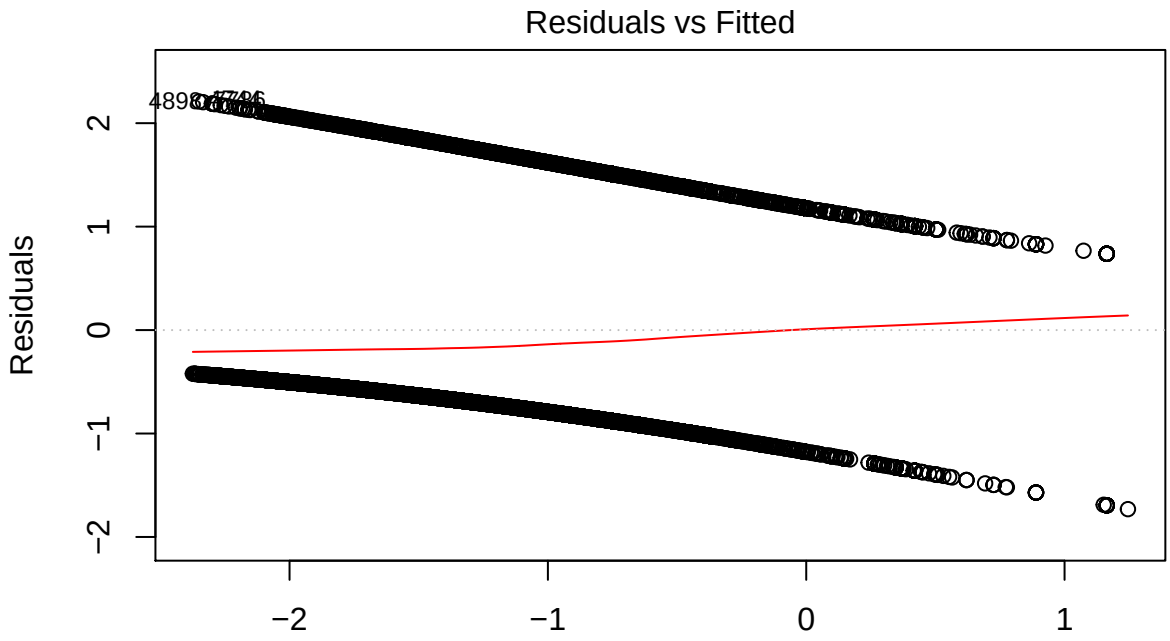
ı(TARGET_FLAG ~ HOME_VAL + PARENT1 + MSTATUS + SEX + EDUCATION + CAR_



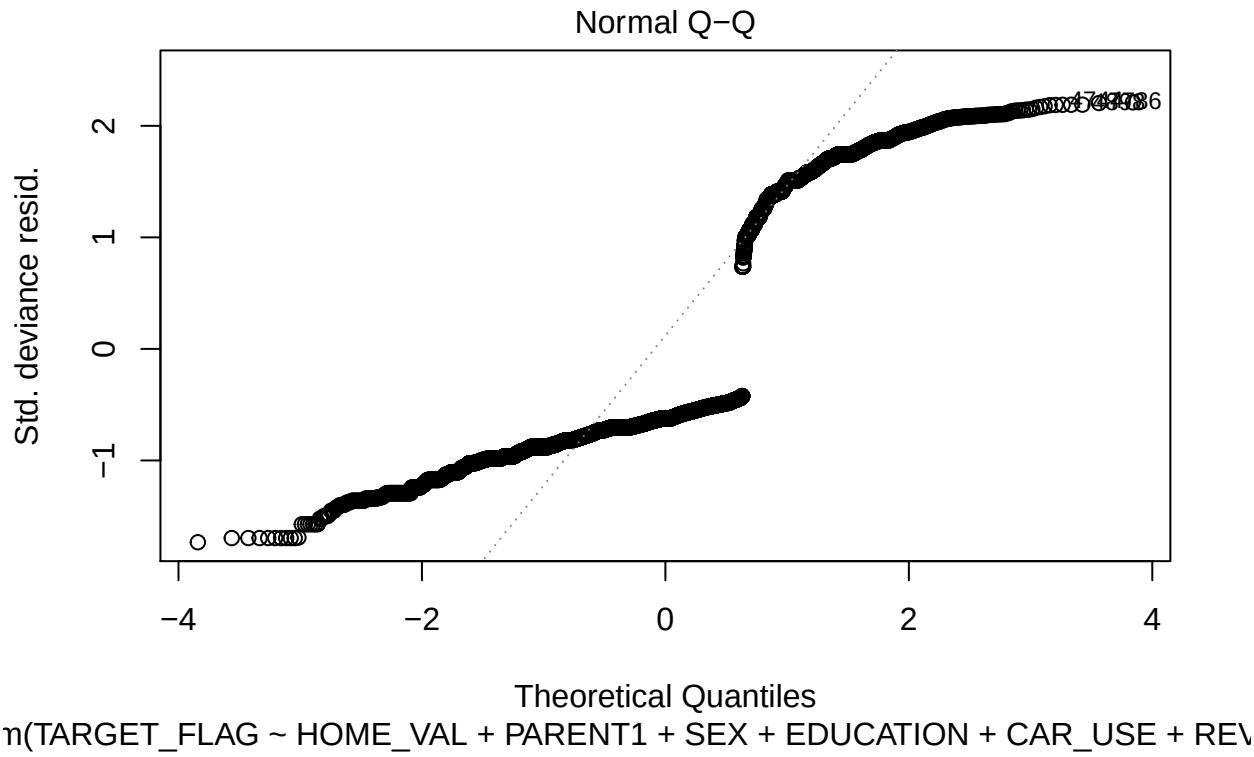
ı(TARGET_FLAG ~ HOME_VAL + PARENT1 + MSTATUS + SEX + EDUCATION + CAR_

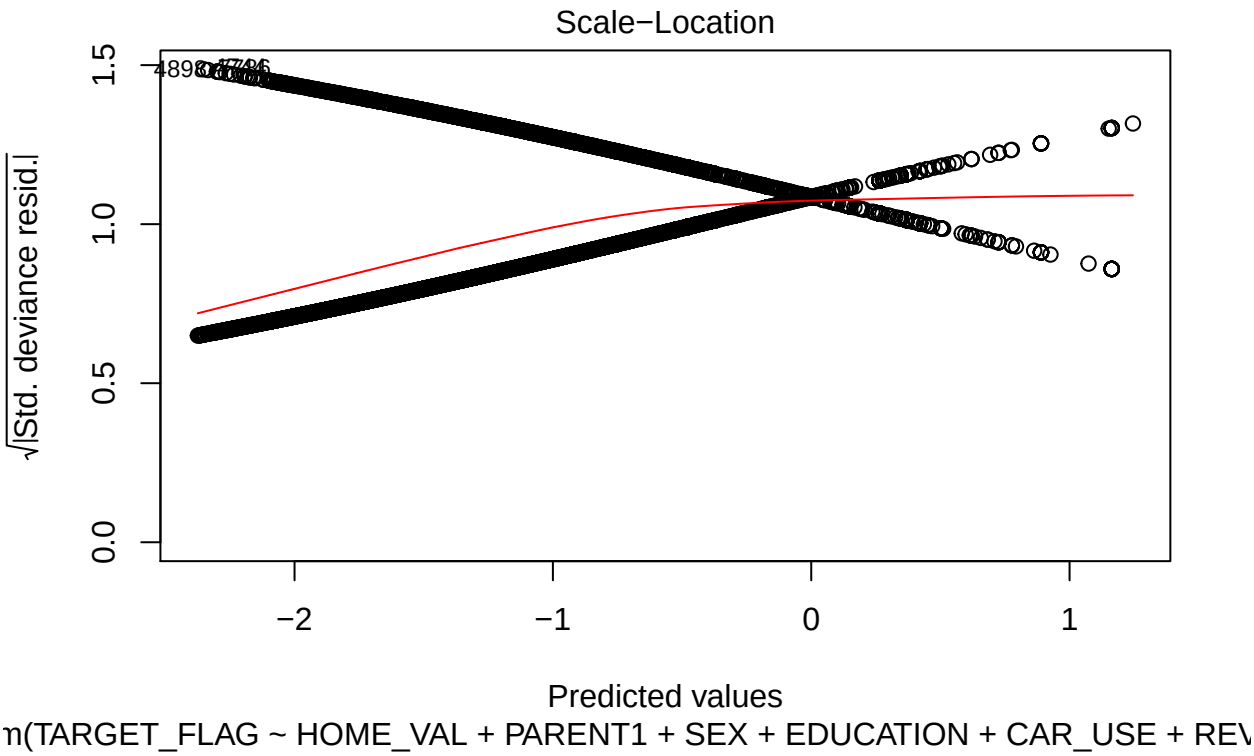
```
##
## Call:
## glm(formula = TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION +
##     CAR_USE + REVOKED, family = "binomial", data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7311  -0.7859  -0.6216   1.0252   2.2098
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.1219412  0.0880653  -1.385    0.166
## HOME_VAL    -0.0001623  0.0000168  -9.662  < 2e-16 ***
## PARENT1      0.7425502  0.0714450  10.393  < 2e-16 ***
## SEX         -0.2738351  0.0555777  -4.927 8.35e-07 ***
## EDUCATION   -0.3559352  0.0721846  -4.931 8.19e-07 ***
## CAR_USE     -0.7940636  0.0564340 -14.071  < 2e-16 ***
## REVOKED      0.8989160  0.0723041  12.432  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9418.0  on 8160  degrees of freedom
## Residual deviance: 8750.4  on 8154  degrees of freedom
## AIC: 8764.4
```

```
##
## Number of Fisher Scoring iterations: 4
```

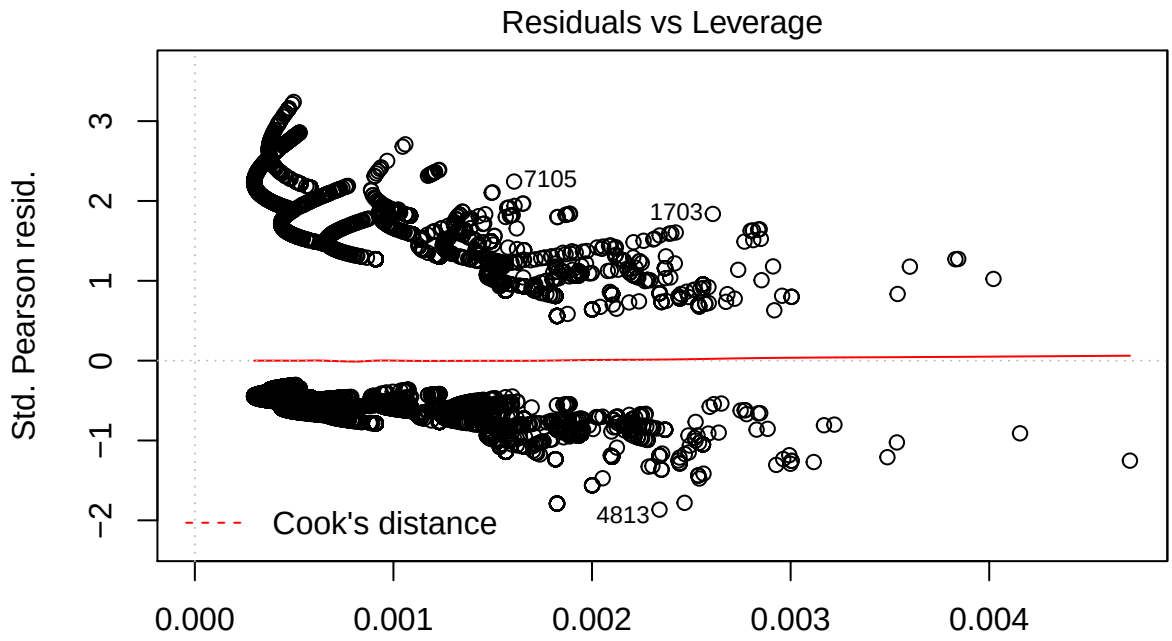


ɳ(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REV


Normal Q−Q
Std. deviance resid.
−1
0
1
2
−4
−2
0
2
4
Theoretical Quantiles
m(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REV



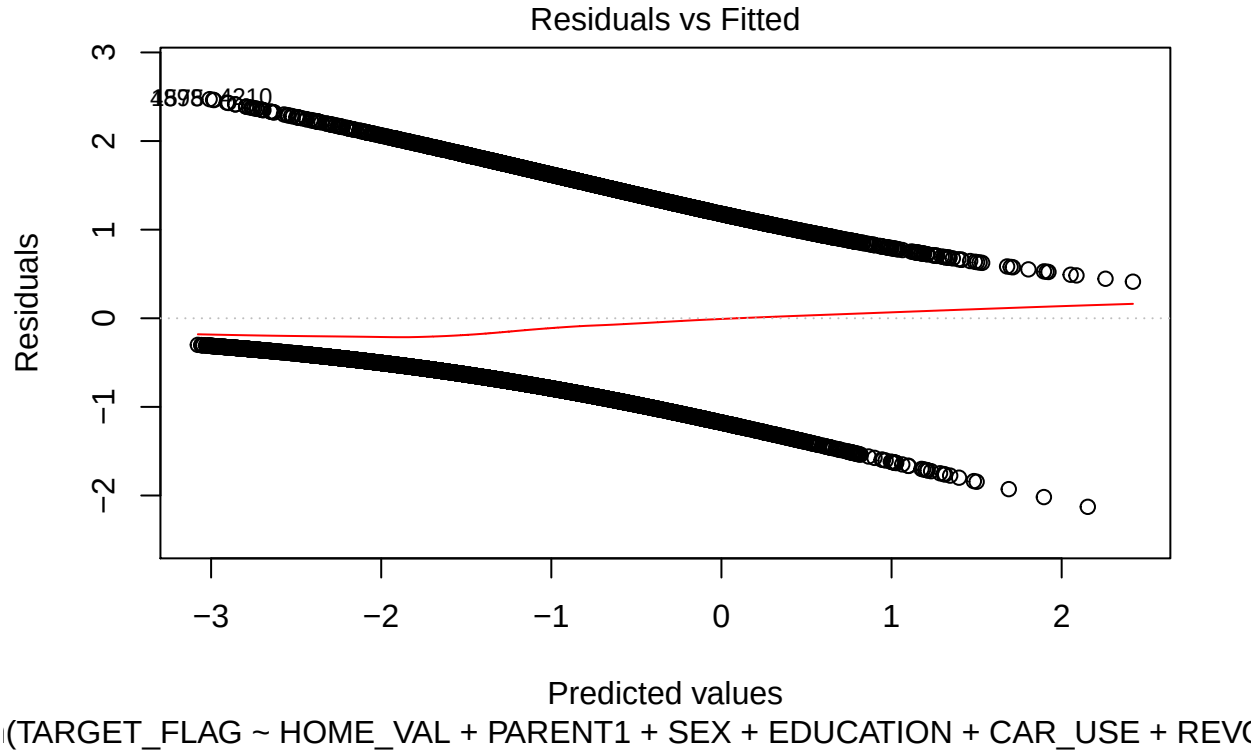
ɳ(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REV

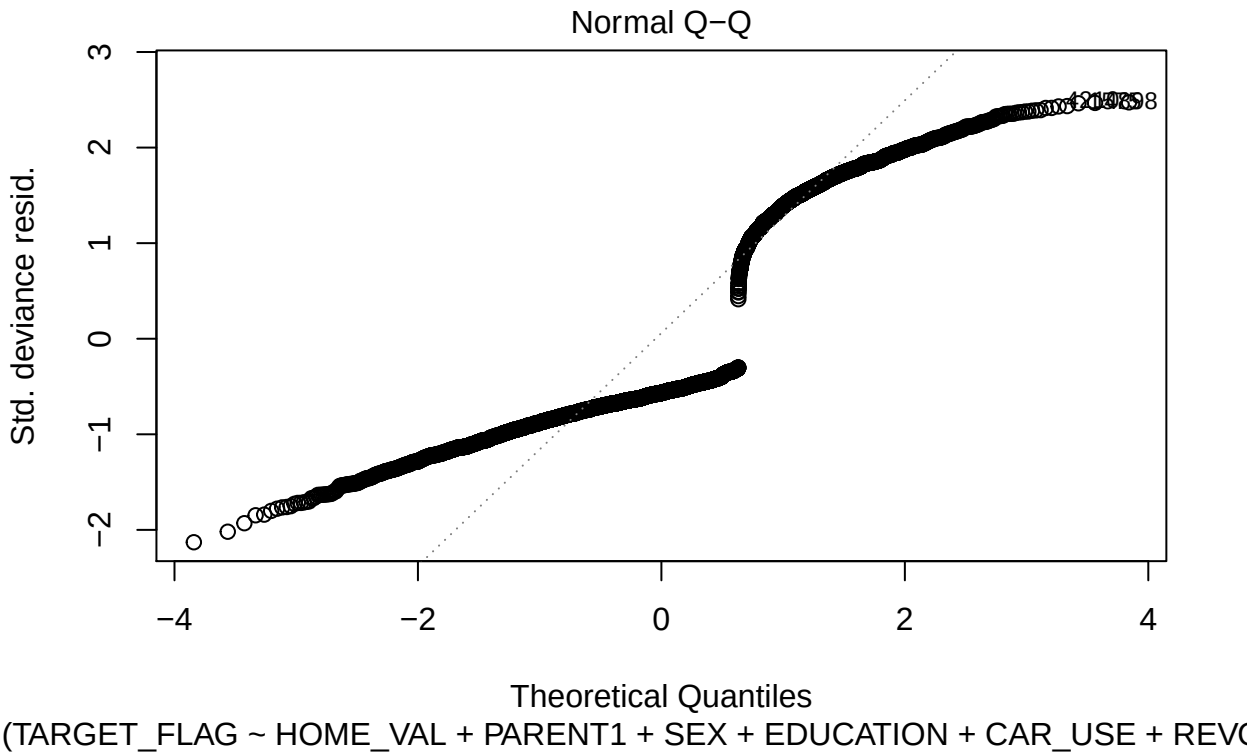


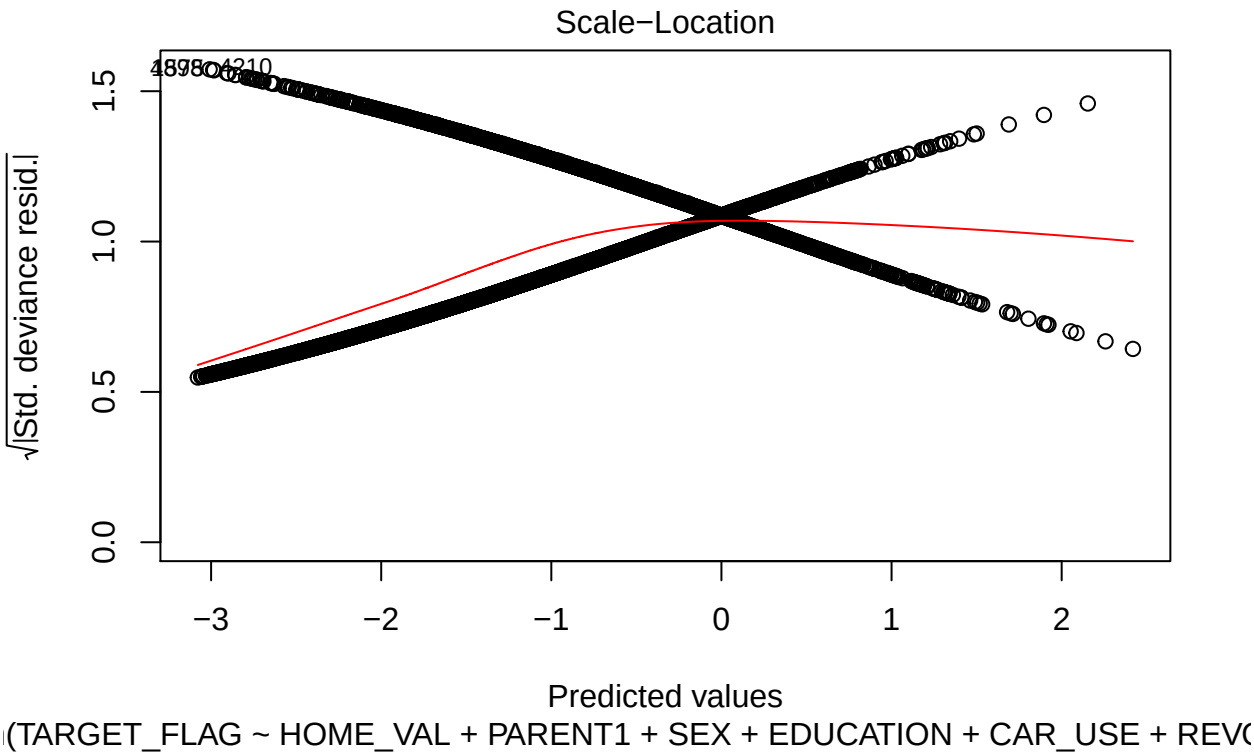
n(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REV

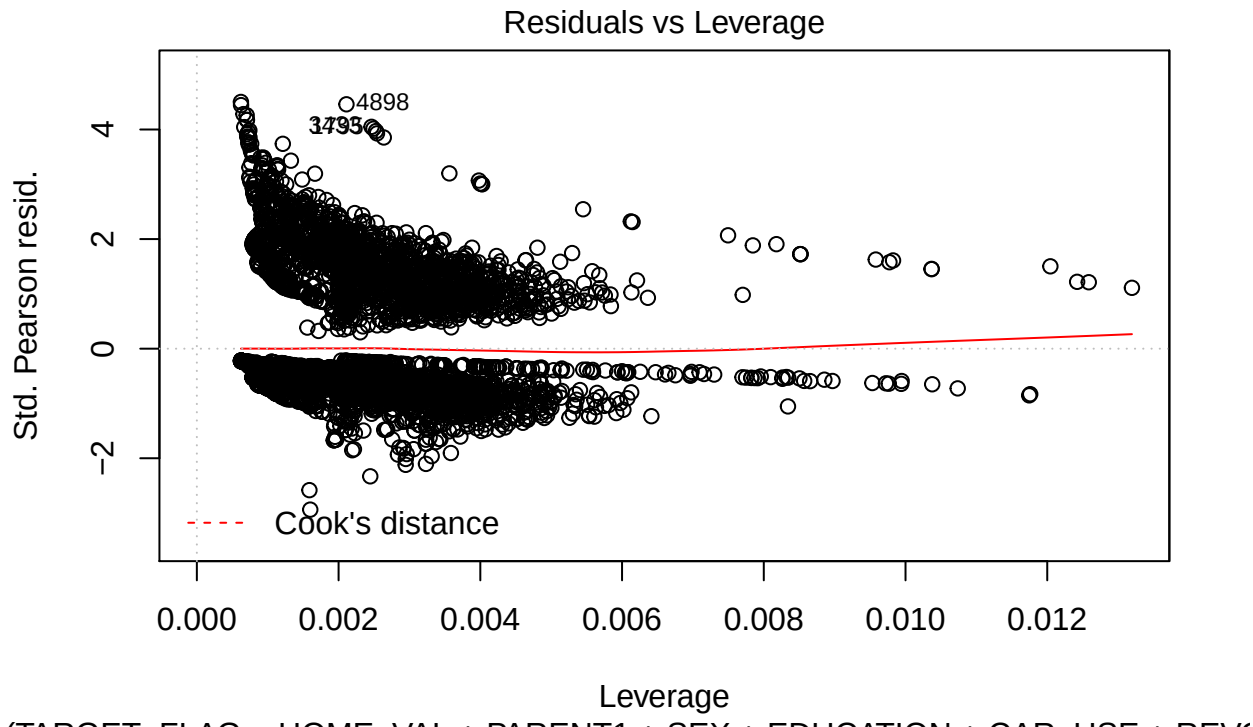
```
##
## Call:
## glm(formula = TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION +
##     CAR_USE + REVOKED + log10(INCOME + 1) + JOB + CLM_FREQ, family = "binomial",
##     data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1277  -0.7613  -0.5665   0.8789   2.4726
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.8238697  0.1754519  -4.696 2.66e-06 ***
## HOME_VAL          -0.0001275  0.0000179  -7.121 1.07e-12 ***
## PARENT1            0.7377840  0.0733318  10.061  < 2e-16 ***
## SEX               -0.1893301  0.0591021  -3.203  0.00136 **
## EDUCATION         -0.1686661  0.0810705  -2.080  0.03748 *
## CAR_USE           -0.6464056  0.0739799  -8.738  < 2e-16 ***
## REVOKED            0.8769533  0.0744377  11.781  < 2e-16 ***
## log10(INCOME + 1) -0.0562881  0.0280631  -2.006  0.04488 *
## JOBClerical        0.3639580  0.1352309   2.691  0.00712 **
## JOBDoctor         -0.3889640  0.2387730  -1.629  0.10331
## JOBHome Maker      0.4333881  0.1593309   2.720  0.00653 **
## JOBLawyer          0.1227946  0.1548887   0.793  0.42790
## JOBManager        -0.4476218  0.1482436  -3.020  0.00253 **
## JOBProfessional    0.1277468  0.1363286   0.937  0.34873
```

```
## JOBStudent          0.3368188  0.1466383   2.297  0.02162 *
## JOBz_Blue Collar    0.3609766  0.1201587   3.004  0.00266 **
## CLM_FREQ            0.3736684  0.0218525  17.100  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9418.0  on 8160  degrees of freedom
## Residual deviance: 8379.6  on 8144  degrees of freedom
## AIC: 8413.6
##
## Number of Fisher Scoring iterations: 4
```




Normal Q−Q
Std. deviance resid.
3
2
1
0
−1
−2
−4
−2
0
2
4
Theoretical Quantiles
(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REVO


Scale−Location
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
−3
−2
−1
0
1
2
Predicted values
(TARGET_FLAG ~ HOME_VAL + PARENT1 + SEX + EDUCATION + CAR_USE + REVO



Select Model

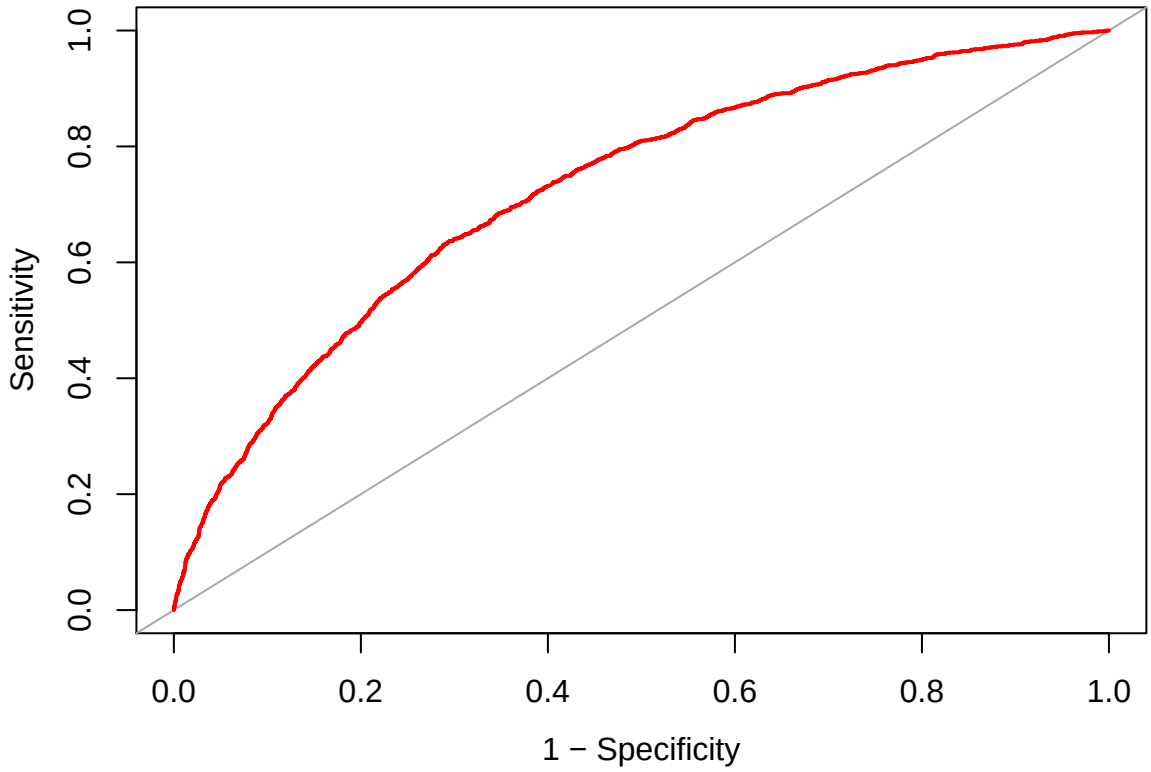
We going to select model number 6 from logistic regression for which gives a better AIC result.in model number 6, we are using log() function on the income variable and selecting the most significant variables.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 5918 1950
##          1   90  203
##
##                Accuracy : 0.75
##                  95% CI : (0.7405, 0.7594)
##     No Information Rate : 0.7362
##     P-Value [Acc > NIR] : 0.002256
##
##                   Kappa : 0.1097
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.98502
##             Specificity : 0.09429
##          Pos Pred Value : 0.75216
##          Neg Pred Value : 0.69283
```

```
##              Prevalence : 0.73618
##          Detection Rate : 0.72516
##    Detection Prevalence : 0.96410
##       Balanced Accuracy : 0.53965
##
##        'Positive' Class : 0
##
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```




```
##   INDEX TARGET_FLAG TARGET_AMT KIDSDRIV AGE HOMEKIDS YOJ   INCOME PARENT1
## 1     3          NA         NA        0  48        0  11 4.062206       0
## 2     9          NA         NA        1  40        1  11 4.048830       1
## 3    10          NA         NA        0  44        2  12 3.988559       1
## 4    18          NA         NA        0  35        2  NA 3.710117       1
## 5    21          NA         NA        0  59        0  12 4.226858       0
## 6    30          NA         NA        0  46        0  14 1.000000       0
##   HOME_VAL MSTATUS SEX EDUCATION           JOB TRAVTIME CAR_USE BLUEBOOK TIF
## 1        2       0   1         1       Manager       26       1      703   1
## 2        2       0   1         1       Manager       21       1      540   6
## 3        2       0   0         1 z_Blue Collar       30       0     1189  10
## 4        2       0   1         1      Clerical       74       1     1373   6
## 5        2       0   1         1       Manager       45       1      345   1
```

```
## 6      636       1   1         1  Professional        7       0      864   1
##      CAR_TYPE RED_CAR OLDCLAIM CLM_FREQ REVOKED MVR_PTS CAR_AGE URBANICITY
## 1         Van       1        1        0       0       2      10           1
## 2     Minivan       0      272        1       0       2       1           1
## 3       z_SUV       0        1        0       0       0      10           0
## 4      Pickup       0        1        0       1       0       4           0
## 5     Minivan       1      494        2       0       4       1           1
## 6 Panel Truck       0      137        1       0       2      12           1
##          pred target.pred
## 1 -2.31579409           0
## 2 -1.20427696           0
## 3  0.06668222           0
## 4  0.11228532           0
## 5 -1.56923964           0
## 6 -0.77846512           0
```