Regression Modelling - Poisson

Reading the Data

```
library(DMwR)
```

```
## Warning: package 'DMwR' was built under R version 3.2.5
```

```
## Loading required package: lattice
```

```
## Loading required package: grid
```

```
datapath<- "C:/Assignment Data"
Businesses<- read.csv(file = paste(datapath, "Master.final.csv", sep = "/"),header = TRUE)
nrow(Businesses)
```

```
## [1] 28710
```

```
Businesses <- Businesses[,-c(1,5)] #Remove index and city column
```

Putting variables in the right format

```
Businesses$Business <- as.character(Businesses$Business)
Businesses$Address <- as.character(Businesses$Address)
Businesses$Under.5.Years <- as.integer(Businesses$Under.5.Years)
Businesses$Start.Date <- as.Date(Businesses$Start.Date,format = "%m/%d/%y")
Businesses$End.Date <- as.Date(Businesses$End.Date, format = "%m/%d/%y")
```

Consider only subset of businesses that have start date after 2010

```
start.2010<-Businesses[Businesses$Start.Date >= "2010-01-01",]
nrow(start.2010)
```

```
## [1] 7747
```

```
which(start.2010$Duration<0) # remove rows with negative age values
```

```
## [1]  846 1001 1763 2354 4884 5208 5323 5406 7051
```

```
start.2010<- start.2010[-c(846,1001,1763,2354,4884,5208, 5323, 5406, 7051),]
```

Scale data

```
scaled.start.2010<- as.data.frame(scale(start.2010[,c(12:75)]))
Businesses.2010<- as.data.frame(cbind(start.2010[,c(1:11)], scaled.start.2010))
```

Test and holdout

```
set.seed(123)
l<- 1:nrow(Businesses.2010)
train<- sample(l, 6000, replace = FALSE)
data.train<- Businesses.2010[train,]
data.hold<- Businesses.2010[-train,]
all.train<- as.data.frame(cbind(Duration=data.train$Duration, data.train[,12:75]))
all.hold<- as.data.frame(cbind(Duration=data.hold$Duration, data.hold[,12:75]))
head(all.train)
```

```
##       Duration Total.Population       Male     Female Under.5.Years
## 8132      2194        1.2927492  1.2898286  1.2749536     1.5843804
## 22691      736        0.5447035  0.5459003  0.5347417     0.4954828
## 11594      759        0.5238941  0.2012091  0.8429815     0.1496185
## 25380     2209       -1.5707015 -1.5259526 -1.5908949    -1.4567488
## 26939      746        0.2488945  0.3585472  0.1336086    -0.2196868
## 1089       722        1.2927492  1.2898286  1.2749536     1.5843804
##       X5.to.9.Years X10.to.14.Years X15.to.17.Years X18.to.24.Years
## 8132      1.9028454       1.5481406       1.6403378      1.38193140
## 22691     0.3918858       0.3372502       0.4911394      0.08553407
## 11594     0.3131520       0.7462346       0.8800292      0.23893535
## 25380    -1.3865253      -1.3715879      -1.3310129     -1.63694313
## 26939    -0.6909076      -0.6948897      -0.7790612     -0.84739615
## 1089      1.9028454       1.5481406       1.6403378      1.38193140
##       X25.to.34.Years X35.to.44.Years X45.to.54.Years X55.to.64.Years
## 8132      0.285637516       1.2862011       1.3428557       0.9981821
## 22691    -0.007061436       0.5772959       1.0607696       1.1848181
## 11594    -0.511761966       0.1011301       0.8209964       1.2736552
## 25380    -0.725283268      -1.2587329      -1.5320219      -1.6213448
## 26939     0.725901724       0.6915282       0.7165645       0.8099877
## 1089      0.285637516       1.2862011       1.3428557       0.9981821
##       X65.to.74.Years X75.to.84.Years X85.Years.and.over White.Alone
## 8132        0.6842248      -0.1648225          0.3390161  0.02623489
## 22691       0.6446565       0.5400815          0.2130821  0.77787009
## 11594       1.8187335       2.6319493          1.4312512 -1.27791931
## 25380      -1.7819857      -1.7693493         -1.7001456 -0.75102372
## 26939       0.6070341       0.7794007          1.4627347  0.51867769
## 1089        0.6842248      -0.1648225          0.3390161  0.02623489
##       Black.or.African.American.Alone
## 8132                       -0.1057849
## 22691                      -0.7619949
## 11594                       2.8269748
## 25380                      -0.8237165
## 26939                      -0.2361156
## 1089                       -0.1057849
##       American.Indian.and.Alaska.Native.Alone Asian.Alone
## 8132                              -0.1717842 -0.65653452
```

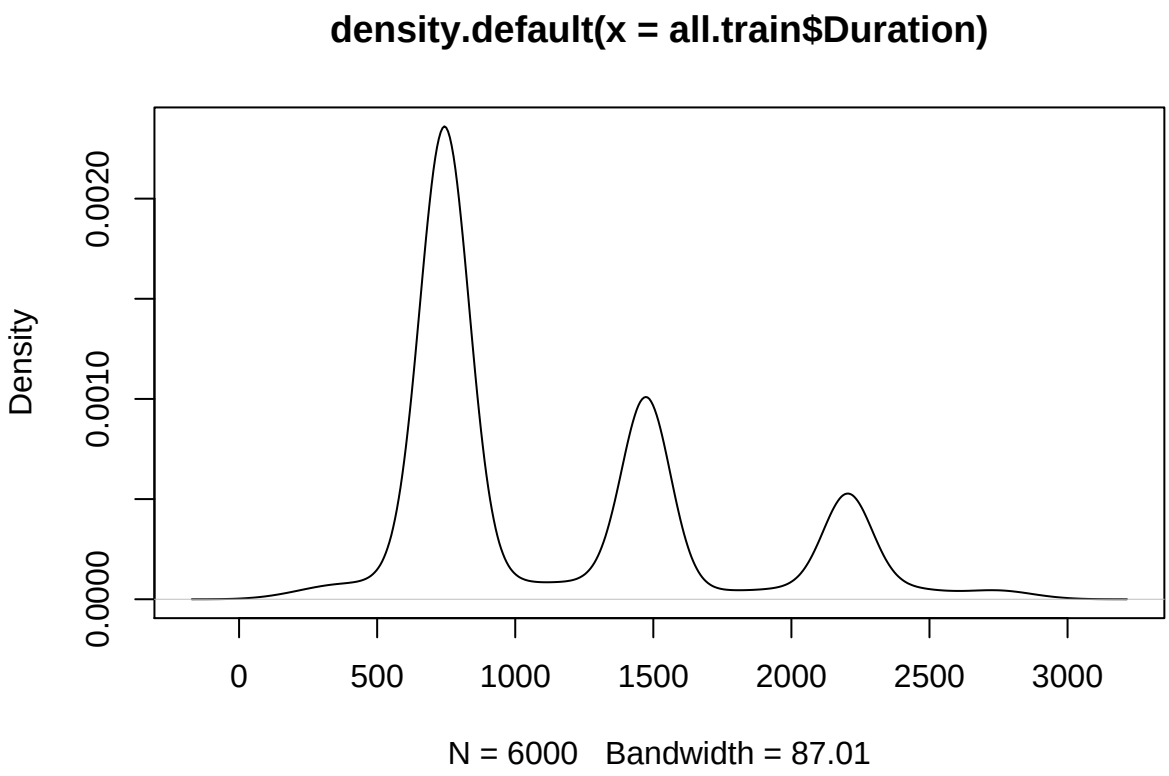
```
## 22691                              -0.2200630  0.02105938
## 11594                               1.5662514 -0.83126165
## 25380                              -0.4476629 -0.37181351
## 26939                               0.6144700  1.13099767
## 1089                               -0.1717842 -0.65653452
##       Native.Hawaiian.and.Other.Pacific.Islander.Alone
## 8132                                         1.3011535
## 22691                                        4.2221161
## 11594                                       -0.2288745
## 25380                                       -0.7388838
## 26939                                        1.3938825
## 1089                                         1.3011535
##       Some.Other.Race.Alone Two.or.More.races Average.Household.Size
## 8132              3.1411867         0.6182136              1.5644988
## 22691             0.8634612         1.6248992              0.6515205
## 11594            -0.6541360        -0.7231588              0.3471944
## 25380            -0.6164870        -1.0237207             -1.6309251
## 26939            -0.5011296         1.6892116             -1.1744359
## 1089              3.1411867         0.6182136              1.5644988
##       Less.Than.High.School High.School.Graduate..includes.equivalency.
## 8132             2.01270339                                    1.3103031
## 22691            0.57002601                                    0.9166360
## 11594            0.02913211                                    1.1574025
## 25380           -1.09773018                                   -1.4040282
## 26939           -0.34218186                                   -0.1654297
## 1089             2.01270339                                    1.3103031
##       Some.college Bachelor.s.degree Master.s.degree
## 8132     0.6644645       -0.75194957      -0.8421556
## 22691    0.8009285       -0.08617848      -0.2223663
## 11594    1.9040125       -0.55465495      -0.5142715
## 25380   -1.5767293       -0.23332411      -0.2983975
## 26939    0.6362619        1.38056161       1.4910489
## 1089     0.6644645       -0.75194957      -0.8421556
##       Professional.school.degree Doctorate.degree Enrolled.In.School
## 8132                   -0.7865857      -1.01392850          1.4612704
## 22691                  -0.5658398      -0.40961890          0.3147150
## 11594                  -0.6931642      -0.81767362          0.5912048
## 25380                   0.3412527       0.04312752         -1.7002411
## 26939                   0.7744571       1.35667508         -0.4835067
## 1089                   -0.7865857      -1.01392850          1.4612704
##       Not.Enrolled.In.School In.labor.force. In.Armed.Forces   Civilian.
## 8132               1.1468472      0.94769486      0.01159068  0.94789432
## 22691              0.6269380      0.69899950      0.40052675  0.69879707
## 11594              0.5223282      0.02954711      3.35644087  0.02656053
## 25380             -1.4627243     -1.24701047     -0.68849425 -1.24667256
## 26939              0.5923986      0.67514250      0.71167560  0.67465732
## 1089               1.1468472      0.94769486      0.01159068  0.94789432
##         Employed   Unemployed Not.In.labor.force Employed.1 Unemployed.1
## 8132   0.8252999  1.108099910          1.0458887  0.8252999  1.108099910
## 22691  0.7076779  0.242039074          0.1684397  0.7076779  0.242039074
## 11594 -0.2916250  1.853946331          1.3701055 -0.2916250  1.853946331
## 25380 -1.1184752 -1.266027058         -1.7385140 -1.1184752 -1.266027058
## 26939  0.7230037  0.003333749          0.1561901  0.7230037  0.003333749
## 1089   0.8252999  1.108099910          1.0458887  0.8252999  1.108099910
```

```
##       Less.than..10.000 X.10.000.to..14.999 X.15.000.to..19.999
## 8132         -0.2715118           0.8946025           0.7370536
## 22691        -0.6485605          -0.2249131           0.1456465
## 11594         1.1227736           1.2868415           1.4528079
## 25380        -1.2613751          -1.4359511          -1.6619363
## 26939         2.4464177           1.4224908           1.2496065
## 1089         -0.2715118           0.8946025           0.7370536
##       X.20.000.to..24.999 X.25.000.to..29.999 X.30.000.to..34.999
## 8132            0.9903007           1.5892973           1.1830218
## 22691           0.2422669           0.6203776           1.0662715
## 11594           1.4945761           1.0217152           0.9286107
## 25380          -1.4625190          -1.8027613          -1.4586713
## 26939           0.4677435           0.8504330           0.7404163
## 1089            0.9903007           1.5892973           1.1830218
##       X.35.000.to..39.999 X.40.000.to..44.999 X.45.000.to..49.999
## 8132            1.4814911           0.9791917           0.7563937
## 22691           0.8255871           0.4064611           0.7563937
## 11594           0.8682024           0.1880817           0.5251519
## 25380          -1.1995626          -1.5795545          -1.3676929
## 26939           0.6829188           0.6104191           0.7492419
## 1089            1.4814911           0.9791917           0.7563937
##       X.50.000.to..59.999 X.60.000.to..74.999 X.75.000.to..99.999
## 8132            0.9804451           0.7009831          0.02350951
## 22691           0.7471548           0.4349165          0.65767733
## 11594           0.6638368           0.1395375         -0.25147569
## 25380          -1.4038374          -1.2009421         -0.71993581
## 26939           0.4749828           1.3920798          1.71193654
## 1089            0.9804451           0.7009831          0.02350951
##       X.100.000.to..124.999 X.125.000.to..149.999 X.150.000.to..199.999
## 8132             -0.3014263            -0.5468837          -0.628460761
## 22691             0.2562143             0.2076192          -0.008825026
## 11594            -0.6283567            -0.4061562          -0.682038287
## 25380            -0.3801520            -0.2569512          -0.005330840
## 26939             0.8182285             0.5857184           0.806485152
## 1089             -0.3014263            -0.5468837          -0.628460761
##       X.200.000.or.More
## 8132         -0.7320642
## 22691        -0.3947752
## 11594        -0.6890458
## 25380         0.5834250
## 26939         0.2991297
## 1089         -0.7320642
##       Median.household.income..In.2014.Inflation.Adjusted.Dollars.
## 8132                                                    -0.6494192
## 22691                                                   -0.2246528
## 11594                                                   -0.9071810
## 25380                                                    1.5972166
## 26939                                                   -0.3240313
## 1089                                                    -0.6494192
##       Average.household.income..In.2014.Inflation.Adjusted.Dollars.
## 8132                                                     -0.7658182
## 22691                                                    -0.3255625
## 11594                                                    -0.8792634
## 25380                                                     1.7143533
```

```
## 26939                                                    -0.2187971
## 1089                                                     -0.7658182
##       With.earnings No.earnings
## 8132     0.45533018  0.02261441
## 22691    0.38277004 -0.15637951
## 11594   -0.04333044  1.99939386
## 25380   -0.97456107 -1.48436728
## 26939    1.11116213  1.80323613
## 1089     0.45533018  0.02261441
##       Per.capita.income..In.2014.Inflation.adjusted.dollars.
## 8132                                              -0.82303362
## 22691                                             -0.48694315
## 11594                                             -0.72180013
## 25380                                              2.34696347
## 26939                                              0.08069033
## 1089                                              -0.82303362
##       Living.in.Poverty At.or.Above.Poverty.Level       Crime
## 8132          0.9846767                1.04511872  0.35631229
## 22691        -0.2823452                0.75792844  0.38827195
## 11594         0.9267543                0.03500355  1.38276459
## 25380        -1.5377341               -1.19096513 -0.42713139
## 26939         0.4345352                0.53265440 -0.05657204
## 1089          0.9846767                1.04511872  0.35631229
```

Distribution of the Age predictor (dependent variable)

```
plot(density(all.train$Duration))
```



Poisson regression using all predictors

```
full.model.train<- glm(Duration~., data = all.train, family = poisson(link = "log"))
summary(full.model.train)
```

```
##
## Call:
## glm(formula = Duration ~ ., family = poisson(link = "log"), data = all.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -39.239  -12.960   -7.943    9.304   48.128
##
## Coefficients: (9 not defined because of singularities)
##                                                                 Estimate
## (Intercept)                                                    7.0492756
## Total.Population                                              -6.9077422
## Male                                                           1.6549026
## Female                                                                NA
## Under.5.Years                                                  0.3242169
## X5.to.9.Years                                                  1.4193198
## X10.to.14.Years                                                0.5436477
## X15.to.17.Years                                               -0.1802287
## X18.to.24.Years                                                0.2004242
```

```
## X25.to.34.Years                                                 1.7864567
## X35.to.44.Years                                                 1.6707051
## X45.to.54.Years                                                 0.8241862
## X55.to.64.Years                                                 0.5037609
## X65.to.74.Years                                                 0.8669277
## X75.to.84.Years                                                 0.9554535
## X85.Years.and.over                                                     NA
## White.Alone                                                     5.0533957
## Black.or.African.American.Alone                                 4.5608194
## American.Indian.and.Alaska.Native.Alone                        -0.0674998
## Asian.Alone                                                     0.8844481
## Native.Hawaiian.and.Other.Pacific.Islander.Alone               -0.0024389
## Some.Other.Race.Alone                                           2.8769434
## Two.or.More.races                                                      NA
## Average.Household.Size                                         -0.1336745
## Less.Than.High.School                                          -3.3642034
## High.School.Graduate..includes.equivalency.                    -3.0301760
## Some.college                                                   -2.4406088
## Bachelor.s.degree                                              -3.3661638
## Master.s.degree                                                -1.9964670
## Professional.school.degree                                     -1.2682395
## Doctorate.degree                                                       NA
## Enrolled.In.School                                             -3.6637980
## Not.Enrolled.In.School                                         -7.5737810
## In.labor.force.                                                 5.9288857
## In.Armed.Forces                                                -0.0494891
## Civilian.                                                              NA
## Employed                                                       -0.3103545
## Unemployed                                                             NA
## Not.In.labor.force                                              2.7085704
## Employed.1                                                             NA
## Unemployed.1                                                           NA
## Less.than..10.000                                               0.0358570
## X.10.000.to..14.999                                            -0.1857930
## X.15.000.to..19.999                                             0.5786699
## X.20.000.to..24.999                                             0.1184195
## X.25.000.to..29.999                                            -0.2156970
## X.30.000.to..34.999                                             0.2369262
## X.35.000.to..39.999                                             0.0409103
## X.40.000.to..44.999                                            -0.0439916
## X.45.000.to..49.999                                             0.1565389
## X.50.000.to..59.999                                            -0.3590335
## X.60.000.to..74.999                                             0.1662931
## X.75.000.to..99.999                                             0.6263520
## X.100.000.to..124.999                                          -0.4489964
## X.125.000.to..149.999                                           0.4144020
## X.150.000.to..199.999                                           0.3823444
## X.200.000.or.More                                               0.1447702
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.   -0.2446872
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.   0.0900025
## With.earnings                                                  -0.1000265
## No.earnings                                                            NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.          0.0680016
## Living.in.Poverty                                               0.5084043
```

```
## At.or.Above.Poverty.Level                                  2.7332312
## Crime                                                     -0.1158623
##                                                           Std. Error
## (Intercept)                                                0.0003821
## Total.Population                                           6.1226329
## Male                                                       0.2478715
## Female                                                            NA
## Under.5.Years                                              0.3538691
## X5.to.9.Years                                              0.4021301
## X10.to.14.Years                                            0.3095860
## X15.to.17.Years                                            0.1160752
## X18.to.24.Years                                            0.0538257
## X25.to.34.Years                                            0.1562768
## X35.to.44.Years                                            0.1309212
## X45.to.54.Years                                            0.0871724
## X55.to.64.Years                                            0.0600559
## X65.to.74.Years                                            0.0215133
## X75.to.84.Years                                            0.0241373
## X85.Years.and.over                                                NA
## White.Alone                                                0.7608582
## Black.or.African.American.Alone                            0.6056713
## American.Indian.and.Alaska.Native.Alone                    0.0344255
## Asian.Alone                                                0.1293231
## Native.Hawaiian.and.Other.Pacific.Islander.Alone           0.0036152
## Some.Other.Race.Alone                                      0.4176036
## Two.or.More.races                                                 NA
## Average.Household.Size                                     0.0195487
## Less.Than.High.School                                      0.0688035
## High.School.Graduate..includes.equivalency.                0.0692158
## Some.college                                               0.0775878
## Bachelor.s.degree                                          0.0987676
## Master.s.degree                                            0.0371110
## Professional.school.degree                                 0.0329530
## Doctorate.degree                                                  NA
## Enrolled.In.School                                         0.5452998
## Not.Enrolled.In.School                                     1.1230583
## In.labor.force.                                            1.9046731
## In.Armed.Forces                                            0.0034280
## Civilian.                                                         NA
## Employed                                                   0.1781422
## Unemployed                                                        NA
## Not.In.labor.force                                         1.0143226
## Employed.1                                                        NA
## Unemployed.1                                                      NA
## Less.than..10.000                                          0.0527917
## X.10.000.to..14.999                                        0.0066394
## X.15.000.to..19.999                                        0.0307767
## X.20.000.to..24.999                                        0.0224317
## X.25.000.to..29.999                                        0.0087175
## X.30.000.to..34.999                                        0.0179109
## X.35.000.to..39.999                                        0.0623258
## X.40.000.to..44.999                                        0.0136732
## X.45.000.to..49.999                                        0.0104859
## X.50.000.to..59.999                                        0.0427346
```

```
## X.60.000.to..74.999                                          0.0678527
## X.75.000.to..99.999                                          0.0189870
## X.100.000.to..124.999                                        0.0361919
## X.125.000.to..149.999                                        0.0222275
## X.150.000.to..199.999                                        0.0502093
## X.200.000.or.More                                            0.0439376
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.  0.0160823
## Average.household.income..In.2014.Inflation.Adjusted.Dollars. 0.0072990
## With.earnings                                                0.4024537
## No.earnings                                                         NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.       0.0080584
## Living.in.Poverty                                            0.0173496
## At.or.Above.Poverty.Level                                    0.1280386
## Crime                                                        0.0005615
##                                                               z value
## (Intercept)                                                18450.409
## Total.Population                                              -1.128
## Male                                                           6.676
## Female                                                            NA
## Under.5.Years                                                  0.916
## X5.to.9.Years                                                  3.530
## X10.to.14.Years                                                1.756
## X15.to.17.Years                                               -1.553
## X18.to.24.Years                                                3.724
## X25.to.34.Years                                               11.431
## X35.to.44.Years                                               12.761
## X45.to.54.Years                                                9.455
## X55.to.64.Years                                                8.388
## X65.to.74.Years                                               40.297
## X75.to.84.Years                                               39.584
## X85.Years.and.over                                                NA
## White.Alone                                                    6.642
## Black.or.African.American.Alone                                7.530
## American.Indian.and.Alaska.Native.Alone                       -1.961
## Asian.Alone                                                    6.839
## Native.Hawaiian.and.Other.Pacific.Islander.Alone              -0.675
## Some.Other.Race.Alone                                          6.889
## Two.or.More.races                                                 NA
## Average.Household.Size                                        -6.838
## Less.Than.High.School                                        -48.896
## High.School.Graduate..includes.equivalency.                  -43.779
## Some.college                                                 -31.456
## Bachelor.s.degree                                            -34.082
## Master.s.degree                                              -53.797
## Professional.school.degree                                   -38.486
## Doctorate.degree                                                  NA
## Enrolled.In.School                                            -6.719
## Not.Enrolled.In.School                                        -6.744
## In.labor.force.                                                3.113
## In.Armed.Forces                                              -14.437
## Civilian.                                                         NA
## Employed                                                      -1.742
## Unemployed                                                        NA
## Not.In.labor.force                                             2.670
```

```
## Employed.1                                                      NA
## Unemployed.1                                                    NA
## Less.than..10.000                                            0.679
## X.10.000.to..14.999                                        -27.983
## X.15.000.to..19.999                                         18.802
## X.20.000.to..24.999                                          5.279
## X.25.000.to..29.999                                        -24.743
## X.30.000.to..34.999                                         13.228
## X.35.000.to..39.999                                          0.656
## X.40.000.to..44.999                                         -3.217
## X.45.000.to..49.999                                         14.929
## X.50.000.to..59.999                                         -8.401
## X.60.000.to..74.999                                          2.451
## X.75.000.to..99.999                                         32.989
## X.100.000.to..124.999                                      -12.406
## X.125.000.to..149.999                                       18.644
## X.150.000.to..199.999                                        7.615
## X.200.000.or.More                                            3.295
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.  -15.215
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.  12.331
## With.earnings                                               -0.249
## No.earnings                                                     NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.       8.439
## Living.in.Poverty                                           29.304
## At.or.Above.Poverty.Level                                   21.347
## Crime                                                     -206.345
##                                                          Pr(>|z|)
## (Intercept)                                               < 2e-16 ***
## Total.Population                                         0.259223
## Male                                                     2.45e-11 ***
## Female                                                         NA
## Under.5.Years                                            0.359559
## X5.to.9.Years                                            0.000416 ***
## X10.to.14.Years                                          0.079080 .
## X15.to.17.Years                                          0.120498
## X18.to.24.Years                                          0.000196 ***
## X25.to.34.Years                                           < 2e-16 ***
## X35.to.44.Years                                           < 2e-16 ***
## X45.to.54.Years                                           < 2e-16 ***
## X55.to.64.Years                                           < 2e-16 ***
## X65.to.74.Years                                           < 2e-16 ***
## X75.to.84.Years                                           < 2e-16 ***
## X85.Years.and.over                                             NA
## White.Alone                                              3.10e-11 ***
## Black.or.African.American.Alone                          5.07e-14 ***
## American.Indian.and.Alaska.Native.Alone                  0.049909 *
## Asian.Alone                                              7.97e-12 ***
## Native.Hawaiian.and.Other.Pacific.Islander.Alone         0.499907
## Some.Other.Race.Alone                                    5.61e-12 ***
## Two.or.More.races                                              NA
## Average.Household.Size                                   8.03e-12 ***
## Less.Than.High.School                                     < 2e-16 ***
## High.School.Graduate..includes.equivalency.               < 2e-16 ***
## Some.college                                              < 2e-16 ***
```

```
## Bachelor.s.degree                                                 < 2e-16 ***
## Master.s.degree                                                   < 2e-16 ***
## Professional.school.degree                                        < 2e-16 ***
## Doctorate.degree                                                      NA
## Enrolled.In.School                                              1.83e-11 ***
## Not.Enrolled.In.School                                          1.54e-11 ***
## In.labor.force.                                                 0.001853 **
## In.Armed.Forces                                                   < 2e-16 ***
## Civilian.                                                             NA
## Employed                                                        0.081478 .
## Unemployed                                                            NA
## Not.In.labor.force                                              0.007578 **
## Employed.1                                                            NA
## Unemployed.1                                                          NA
## Less.than..10.000                                               0.497001
## X.10.000.to..14.999                                               < 2e-16 ***
## X.15.000.to..19.999                                               < 2e-16 ***
## X.20.000.to..24.999                                             1.30e-07 ***
## X.25.000.to..29.999                                               < 2e-16 ***
## X.30.000.to..34.999                                               < 2e-16 ***
## X.35.000.to..39.999                                             0.511571
## X.40.000.to..44.999                                             0.001294 **
## X.45.000.to..49.999                                               < 2e-16 ***
## X.50.000.to..59.999                                               < 2e-16 ***
## X.60.000.to..74.999                                             0.014254 *
## X.75.000.to..99.999                                               < 2e-16 ***
## X.100.000.to..124.999                                             < 2e-16 ***
## X.125.000.to..149.999                                             < 2e-16 ***
## X.150.000.to..199.999                                           2.64e-14 ***
## X.200.000.or.More                                               0.000985 ***
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.      < 2e-16 ***
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.     < 2e-16 ***
## With.earnings                                                   0.803715
## No.earnings                                                           NA
## Per.capita.income..In.2014.Inflation.adjusted.dollars.            < 2e-16 ***
## Living.in.Poverty                                                 < 2e-16 ***
## At.or.Above.Poverty.Level                                         < 2e-16 ***
## Crime                                                             < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1599719  on 5999  degrees of freedom
## Residual deviance: 1504879  on 5944  degrees of freedom
## AIC: 1557663
##
## Number of Fisher Scoring iterations: 4
```

```r
# Calculate R-square of test model
cor(all.train$Duration, predict(full.model.train, type = "link"))^2
```

```
## [1] 0.05475966
```

```r
# Holdout validation
pred.hold<-predict(full.model.train, newdata = all.hold, type = "link")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading
```

```r
# Calculate R-square of holdout model
cor(all.hold$Duration, pred.hold)^2
```

```
## [1] 0.02878323
```

```r
# calculate MSE
mean((all.hold$Duration-pred.hold)^2)
```

```
## [1] 1602988
```

Check for over-dispersion in full model

```r
disp<- sum(residuals(full.model.train, type = "pearson")^2)/full.model.train$df.residual
disp
```

```
## [1] 273.6558
```

```r
# Quasi-poisson model to overcome over-dispersion

full.model.quasi<- glm(Duration~., data = all.train, family = quasipoisson(link = "log"))

# Calculate R-square of test model
cor(all.train$Duration, predict(full.model.quasi, type = "link"))^2
```

```
## [1] 0.05475966
```

```r
# Holdout validation
Predict.hold.quasi<- predict(full.model.quasi, newdata= all.hold, type= "link")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading
```

```r
# Calculate R-square of test model
cor(all.hold$Duration, Predict.hold.quasi)^2
```

```
## [1] 0.02878323
```

```r
# calculate MSE
mean((all.hold$Duration-Predict.hold.quasi)^2)
```

```
## [1] 1602988
```

```r
#Calculate Accuracy
accuracy.quasi <- table(Predict.hold.quasi, all.hold[,"Duration"])
sum(diag(accuracy.quasi))/sum(accuracy.quasi)
```

```
## [1] 0.001726122
```

stepAIC to identify non-collinear predictors

```r
library(MASS)
# Use stepAIC to identify the non-collienar predictors
step.model<- stepAIC(full.model.train, criterion = "AIC", trace = FALSE)
summary(step.model)
```

```
##
## Call:
## glm(formula = Duration ~ Total.Population + Male + Under.5.Years +
##     X5.to.9.Years + X10.to.14.Years + X15.to.17.Years + X18.to.24.Years +
##     X25.to.34.Years + X35.to.44.Years + X45.to.54.Years + X55.to.64.Years +
##     X65.to.74.Years + X75.to.84.Years + White.Alone + Black.or.African.American.Alone +
##     American.Indian.and.Alaska.Native.Alone + Asian.Alone + Native.Hawaiian.and.Other.Pacific.Islande
##     Some.Other.Race.Alone + Average.Household.Size + Less.Than.High.School +
##     High.School.Graduate..includes.equivalency. + Some.college +
##     Bachelor.s.degree + Master.s.degree + Professional.school.degree +
##     Enrolled.In.School + Not.Enrolled.In.School + In.labor.force. +
##     In.Armed.Forces + Employed + Not.In.labor.force + Less.than..10.000 +
##     X.10.000.to..14.999 + X.15.000.to..19.999 + X.20.000.to..24.999 +
##     X.25.000.to..29.999 + X.30.000.to..34.999 + X.35.000.to..39.999 +
##     X.40.000.to..44.999 + X.45.000.to..49.999 + X.50.000.to..59.999 +
##     X.60.000.to..74.999 + X.75.000.to..99.999 + X.100.000.to..124.999 +
##     X.125.000.to..149.999 + X.150.000.to..199.999 + X.200.000.or.More +
##     Median.household.income..In.2014.Inflation.Adjusted.Dollars. +
##     Average.household.income..In.2014.Inflation.Adjusted.Dollars. +
##     Per.capita.income..In.2014.Inflation.adjusted.dollars. +
##     Living.in.Poverty + At.or.Above.Poverty.Level + Crime, family = poisson(link = "log"),
##     data = all.train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -39.242  -12.960   -7.942    9.304   48.133
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                     7.0492767
## Total.Population                                               -8.4011039
## Male                                                            1.5962496
## Under.5.Years                                                   0.4093096
## X5.to.9.Years                                                   1.5150014
## X10.to.14.Years                                                 0.6168404
## X15.to.17.Years                                                -0.1530619
## X18.to.24.Years                                                 0.2080720
## X25.to.34.Years                                                 1.8203616
```

```
## X35.to.44.Years                                                 1.7005057
## X45.to.54.Years                                                 0.8435699
## X55.to.64.Years                                                 0.5167710
## X65.to.74.Years                                                 0.8655350
## X75.to.84.Years                                                 0.9600121
## White.Alone                                                     5.2354259
## Black.or.African.American.Alone                                 4.7042129
## American.Indian.and.Alaska.Native.Alone                        -0.0759811
## Asian.Alone                                                     0.9153032
## Native.Hawaiian.and.Other.Pacific.Islander.Alone               -0.0031531
## Some.Other.Race.Alone                                           2.9770775
## Average.Household.Size                                         -0.1291200
## Less.Than.High.School                                          -3.3583645
## High.School.Graduate..includes.equivalency.                    -3.0228305
## Some.college                                                   -2.4258844
## Bachelor.s.degree                                              -3.3525250
## Master.s.degree                                                -1.9928677
## Professional.school.degree                                     -1.2622475
## Enrolled.In.School                                             -3.5367637
## Not.Enrolled.In.School                                         -7.3148167
## In.labor.force.                                                 6.3772149
## In.Armed.Forces                                                -0.0486932
## Employed                                                       -0.3493018
## Not.In.labor.force                                              2.9485174
## Less.than..10.000                                               0.0231089
## X.10.000.to..14.999                                            -0.1847366
## X.15.000.to..19.999                                             0.5846608
## X.20.000.to..24.999                                             0.1132018
## X.25.000.to..29.999                                            -0.2169788
## X.30.000.to..34.999                                             0.2403176
## X.35.000.to..39.999                                             0.0256556
## X.40.000.to..44.999                                            -0.0410566
## X.45.000.to..49.999                                             0.1581965
## X.50.000.to..59.999                                            -0.3693189
## X.60.000.to..74.999                                             0.1496201
## X.75.000.to..99.999                                             0.6226263
## X.100.000.to..124.999                                          -0.4568286
## X.125.000.to..149.999                                           0.4100010
## X.150.000.to..199.999                                           0.3701333
## X.200.000.or.More                                               0.1342634
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.   -0.2422426
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.   0.0887704
## Per.capita.income..In.2014.Inflation.adjusted.dollars.          0.0676273
## Living.in.Poverty                                               0.5060927
## At.or.Above.Poverty.Level                                       2.7104248
## Crime                                                          -0.1158590
##                                                                Std. Error
## (Intercept)                                                     0.0003820
## Total.Population                                                1.1770263
## Male                                                            0.0758139
## Under.5.Years                                                   0.0895236
## X5.to.9.Years                                                   0.1162447
## X10.to.14.Years                                                 0.0955312
## X15.to.17.Years                                                 0.0390659
```

```
## X18.to.24.Years                                                0.0441604
## X25.to.34.Years                                                0.0762288
## X35.to.44.Years                                                0.0525661
## X45.to.54.Years                                                0.0389419
## X55.to.64.Years                                                0.0294313
## X65.to.74.Years                                                0.0207697
## X75.to.84.Years                                                0.0156898
## White.Alone                                                    0.2060578
## Black.or.African.American.Alone                                0.1842431
## American.Indian.and.Alaska.Native.Alone                        0.0045475
## Asian.Alone                                                    0.0362111
## Native.Hawaiian.and.Other.Pacific.Islander.Alone               0.0021938
## Some.Other.Race.Alone                                          0.1098271
## Average.Household.Size                                         0.0068041
## Less.Than.High.School                                          0.0646655
## High.School.Graduate..includes.equivalency.                    0.0625848
## Some.college                                                   0.0500952
## Bachelor.s.degree                                              0.0821134
## Master.s.degree                                                0.0341663
## Professional.school.degree                                     0.0224642
## Enrolled.In.School                                             0.1899791
## Not.Enrolled.In.School                                         0.4189952
## In.labor.force.                                                0.6117037
## In.Armed.Forces                                                0.0012240
## Employed                                                       0.0847364
## Not.In.labor.force                                             0.3112562
## Less.than..10.000                                              0.0124877
## X.10.000.to..14.999                                            0.0050999
## X.15.000.to..19.999                                            0.0191313
## X.20.000.to..24.999                                            0.0079026
## X.25.000.to..29.999                                            0.0070285
## X.30.000.to..34.999                                            0.0116024
## X.35.000.to..39.999                                            0.0108282
## X.40.000.to..44.999                                            0.0068927
## X.45.000.to..49.999                                            0.0080893
## X.50.000.to..59.999                                            0.0106573
## X.60.000.to..74.999                                            0.0101986
## X.75.000.to..99.999                                            0.0116530
## X.100.000.to..124.999                                          0.0177953
## X.125.000.to..149.999                                          0.0134336
## X.150.000.to..199.999                                          0.0103538
## X.200.000.or.More                                              0.0119824
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.   0.0127225
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.  0.0053572
## Per.capita.income..In.2014.Inflation.adjusted.dollars.         0.0079153
## Living.in.Poverty                                              0.0146435
## At.or.Above.Poverty.Level                                      0.0892828
## Crime                                                          0.0005613
##                                                                 z value
## (Intercept)                                                   18451.670
## Total.Population                                                 -7.138
## Male                                                             21.055
## Under.5.Years                                                     4.572
## X5.to.9.Years                                                    13.033
```

```
## X10.to.14.Years                                                  6.457
## X15.to.17.Years                                                 -3.918
## X18.to.24.Years                                                  4.712
## X25.to.34.Years                                                 23.880
## X35.to.44.Years                                                 32.350
## X45.to.54.Years                                                 21.662
## X55.to.64.Years                                                 17.559
## X65.to.74.Years                                                 41.673
## X75.to.84.Years                                                 61.187
## White.Alone                                                     25.408
## Black.or.African.American.Alone                                 25.533
## American.Indian.and.Alaska.Native.Alone                        -16.708
## Asian.Alone                                                     25.277
## Native.Hawaiian.and.Other.Pacific.Islander.Alone                -1.437
## Some.Other.Race.Alone                                           27.107
## Average.Household.Size                                         -18.977
## Less.Than.High.School                                          -51.934
## High.School.Graduate..includes.equivalency.                    -48.300
## Some.college                                                   -48.426
## Bachelor.s.degree                                              -40.828
## Master.s.degree                                                -58.329
## Professional.school.degree                                     -56.189
## Enrolled.In.School                                             -18.617
## Not.Enrolled.In.School                                         -17.458
## In.labor.force.                                                 10.425
## In.Armed.Forces                                                -39.783
## Employed                                                        -4.122
## Not.In.labor.force                                               9.473
## Less.than..10.000                                                1.851
## X.10.000.to..14.999                                            -36.224
## X.15.000.to..19.999                                             30.560
## X.20.000.to..24.999                                             14.325
## X.25.000.to..29.999                                            -30.871
## X.30.000.to..34.999                                             20.713
## X.35.000.to..39.999                                              2.369
## X.40.000.to..44.999                                             -5.957
## X.45.000.to..49.999                                             19.556
## X.50.000.to..59.999                                            -34.654
## X.60.000.to..74.999                                             14.671
## X.75.000.to..99.999                                             53.431
## X.100.000.to..124.999                                          -25.671
## X.125.000.to..149.999                                           30.521
## X.150.000.to..199.999                                           35.748
## X.200.000.or.More                                               11.205
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.    -19.041
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.    16.570
## Per.capita.income..In.2014.Inflation.adjusted.dollars.           8.544
## Living.in.Poverty                                               34.561
## At.or.Above.Poverty.Level                                       30.358
## Crime                                                         -206.396
##                                                              Pr(>|z|)
## (Intercept)                                                   < 2e-16 ***
## Total.Population                                             9.50e-13 ***
## Male                                                          < 2e-16 ***
```

```
## Under.5.Years                                                  4.83e-06 ***
## X5.to.9.Years                                                   < 2e-16 ***
## X10.to.14.Years                                                1.07e-10 ***
## X15.to.17.Years                                                8.93e-05 ***
## X18.to.24.Years                                                2.46e-06 ***
## X25.to.34.Years                                                 < 2e-16 ***
## X35.to.44.Years                                                 < 2e-16 ***
## X45.to.54.Years                                                 < 2e-16 ***
## X55.to.64.Years                                                 < 2e-16 ***
## X65.to.74.Years                                                 < 2e-16 ***
## X75.to.84.Years                                                 < 2e-16 ***
## White.Alone                                                     < 2e-16 ***
## Black.or.African.American.Alone                                 < 2e-16 ***
## American.Indian.and.Alaska.Native.Alone                         < 2e-16 ***
## Asian.Alone                                                     < 2e-16 ***
## Native.Hawaiian.and.Other.Pacific.Islander.Alone                 0.1506
## Some.Other.Race.Alone                                           < 2e-16 ***
## Average.Household.Size                                          < 2e-16 ***
## Less.Than.High.School                                           < 2e-16 ***
## High.School.Graduate..includes.equivalency.                     < 2e-16 ***
## Some.college                                                    < 2e-16 ***
## Bachelor.s.degree                                               < 2e-16 ***
## Master.s.degree                                                 < 2e-16 ***
## Professional.school.degree                                      < 2e-16 ***
## Enrolled.In.School                                              < 2e-16 ***
## Not.Enrolled.In.School                                          < 2e-16 ***
## In.labor.force.                                                 < 2e-16 ***
## In.Armed.Forces                                                 < 2e-16 ***
## Employed                                                       3.75e-05 ***
## Not.In.labor.force                                              < 2e-16 ***
## Less.than..10.000                                                0.0642 .
## X.10.000.to..14.999                                             < 2e-16 ***
## X.15.000.to..19.999                                             < 2e-16 ***
## X.20.000.to..24.999                                             < 2e-16 ***
## X.25.000.to..29.999                                             < 2e-16 ***
## X.30.000.to..34.999                                             < 2e-16 ***
## X.35.000.to..39.999                                              0.0178 *
## X.40.000.to..44.999                                            2.58e-09 ***
## X.45.000.to..49.999                                             < 2e-16 ***
## X.50.000.to..59.999                                             < 2e-16 ***
## X.60.000.to..74.999                                             < 2e-16 ***
## X.75.000.to..99.999                                             < 2e-16 ***
## X.100.000.to..124.999                                           < 2e-16 ***
## X.125.000.to..149.999                                           < 2e-16 ***
## X.150.000.to..199.999                                           < 2e-16 ***
## X.200.000.or.More                                               < 2e-16 ***
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.    < 2e-16 ***
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.   < 2e-16 ***
## Per.capita.income..In.2014.Inflation.adjusted.dollars.          < 2e-16 ***
## Living.in.Poverty                                               < 2e-16 ***
## At.or.Above.Poverty.Level                                       < 2e-16 ***
## Crime                                                           < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1599719  on 5999  degrees of freedom
## Residual deviance: 1504879  on 5945  degrees of freedom
## AIC: 1557661
##
## Number of Fisher Scoring iterations: 4
```

```
#Calculate R-square of test model
cor(all.train$Duration, predict(step.model, type = "link"))^2
```

```
## [1] 0.05475967
```

```
# holdout Validation
step.pred<-predict(step.model, newdata = all.hold, type = "link")

#Calculate R-square of holdout
cor(all.hold$Duration, step.pred)^2
```

```
## [1] 0.02877809
```

```
# calculate MSE
mean((all.hold$Duration-step.pred)^2)
```

```
## [1] 1602988
```

OLS Regression using log dependent

```
lm.model<- lm(log(Duration)~., data = all.train)

#Calculate R-square of test model
cor(all.train$Duration, predict(lm.model))^2
```

```
## [1] 0.05426696
```

```
# holdout Validation
lm.pred<-predict(lm.model, newdata = all.hold)
```

```
## Warning in predict.lm(lm.model, newdata = all.hold): prediction from a
## rank-deficient fit may be misleading
```

```
#Calculate R-square of holdout
cor(all.hold$Duration, lm.pred)^2
```

```
## [1] 0.02873291
```

```r
#Use step stepwise
lm.step<-stepAIC(lm.model, direction= "both", trace = FALSE)
summary(lm.step)
```

```
##
## Call:
## lm(formula = log(Duration) ~ Male + Under.5.Years + X10.to.14.Years +
##     X15.to.17.Years + X25.to.34.Years + X35.to.44.Years + X45.to.54.Years +
##     X55.to.64.Years + X65.to.74.Years + X75.to.84.Years + White.Alone +
##     Black.or.African.American.Alone + Asian.Alone + Some.Other.Race.Alone +
##     Average.Household.Size + Less.Than.High.School + High.School.Graduate..includes.equivalency. +
##     Some.college + Bachelor.s.degree + Master.s.degree + Professional.school.degree +
##     Enrolled.In.School + Not.Enrolled.In.School + In.Armed.Forces +
##     Not.In.labor.force + X.10.000.to..14.999 + X.15.000.to..19.999 +
##     X.25.000.to..29.999 + X.30.000.to..34.999 + X.40.000.to..44.999 +
##     X.45.000.to..49.999 + X.50.000.to..59.999 + X.60.000.to..74.999 +
##     X.75.000.to..99.999 + X.100.000.to..124.999 + X.125.000.to..149.999 +
##     X.150.000.to..199.999 + Median.household.income..In.2014.Inflation.Adjusted.Dollars. +
##     Average.household.income..In.2014.Inflation.Adjusted.Dollars. +
##     Per.capita.income..In.2014.Inflation.adjusted.dollars. +
##     Living.in.Poverty + At.or.Above.Poverty.Level + Crime, data = all.train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2674 -0.3309 -0.1723  0.3625  1.1644
##
## Coefficients:
##                                                    Estimate
## (Intercept)                                        6.941102
## Male                                               1.206988
## Under.5.Years                                     -0.380918
## X10.to.14.Years                                   -0.464617
## X15.to.17.Years                                   -0.766264
## X25.to.34.Years                                    1.751163
## X35.to.44.Years                                    1.599061
## X45.to.54.Years                                    0.747665
## X55.to.64.Years                                    0.488383
## X65.to.74.Years                                    0.818703
## X75.to.84.Years                                    0.971223
## White.Alone                                        6.890631
## Black.or.African.American.Alone                    6.212758
## Asian.Alone                                        1.199452
## Some.Other.Race.Alone                              3.937791
## Average.Household.Size                            -0.096746
## Less.Than.High.School                             -3.814918
## High.School.Graduate..includes.equivalency.       -3.468612
## Some.college                                      -2.762940
## Bachelor.s.degree                                 -3.937839
## Master.s.degree                                   -2.164243
## Professional.school.degree                        -1.453740
## Enrolled.In.School                                -2.475619
## Not.Enrolled.In.School                            -4.558069
## In.Armed.Forces                                   -0.065993
```

```
## Not.In.labor.force                                                -0.339013
## X.10.000.to..14.999                                               -0.191523
## X.15.000.to..19.999                                                0.627834
## X.25.000.to..29.999                                               -0.227867
## X.30.000.to..34.999                                                0.212095
## X.40.000.to..44.999                                               -0.126657
## X.45.000.to..49.999                                                0.176726
## X.50.000.to..59.999                                               -0.360467
## X.60.000.to..74.999                                                0.244694
## X.75.000.to..99.999                                                0.608399
## X.100.000.to..124.999                                             -0.523397
## X.125.000.to..149.999                                              0.401391
## X.150.000.to..199.999                                              0.476101
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.       -0.330556
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.       0.113267
## Per.capita.income..In.2014.Inflation.adjusted.dollars.              0.127841
## Living.in.Poverty                                                  0.459257
## At.or.Above.Poverty.Level                                          2.681815
## Crime                                                             -0.126293
##                                                                   Std. Error
## (Intercept)                                                         0.005971
## Male                                                                0.371718
## Under.5.Years                                                       0.091785
## X10.to.14.Years                                                     0.173263
## X15.to.17.Years                                                     0.144569
## X25.to.34.Years                                                     0.377645
## X35.to.44.Years                                                     0.279708
## X45.to.54.Years                                                     0.207688
## X55.to.64.Years                                                     0.142116
## X65.to.74.Years                                                     0.116264
## X75.to.84.Years                                                     0.123790
## White.Alone                                                         1.188830
## Black.or.African.American.Alone                                     1.035630
## Asian.Alone                                                         0.209907
## Some.Other.Race.Alone                                               0.673919
## Average.Household.Size                                              0.055796
## Less.Than.High.School                                               0.520835
## High.School.Graduate..includes.equivalency.                         0.486765
## Some.college                                                        0.373311
## Bachelor.s.degree                                                   0.577447
## Master.s.degree                                                     0.276216
## Professional.school.degree                                          0.187077
## Enrolled.In.School                                                  0.443513
## Not.Enrolled.In.School                                              0.984721
## In.Armed.Forces                                                     0.012306
## Not.In.labor.force                                                  0.140842
## X.10.000.to..14.999                                                 0.051478
## X.15.000.to..19.999                                                 0.105261
## X.25.000.to..29.999                                                 0.057243
## X.30.000.to..34.999                                                 0.078099
## X.40.000.to..44.999                                                 0.045252
## X.45.000.to..49.999                                                 0.049656
## X.50.000.to..59.999                                                 0.070006
## X.60.000.to..74.999                                                 0.069872
```

```
## X.75.000.to..99.999                                              0.105182
## X.100.000.to..124.999                                            0.091660
## X.125.000.to..149.999                                            0.095566
## X.150.000.to..199.999                                            0.077498
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.     0.066962
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.    0.058135
## Per.capita.income..In.2014.Inflation.adjusted.dollars.           0.056658
## Living.in.Poverty                                                0.117123
## At.or.Above.Poverty.Level                                        0.454759
## Crime                                                            0.008888
##                                                                  t value
## (Intercept)                                                     1162.546
## Male                                                               3.247
## Under.5.Years                                                     -4.150
## X10.to.14.Years                                                   -2.682
## X15.to.17.Years                                                   -5.300
## X25.to.34.Years                                                    4.637
## X35.to.44.Years                                                    5.717
## X45.to.54.Years                                                    3.600
## X55.to.64.Years                                                    3.437
## X65.to.74.Years                                                    7.042
## X75.to.84.Years                                                    7.846
## White.Alone                                                        5.796
## Black.or.African.American.Alone                                    5.999
## Asian.Alone                                                        5.714
## Some.Other.Race.Alone                                              5.843
## Average.Household.Size                                            -1.734
## Less.Than.High.School                                             -7.325
## High.School.Graduate..includes.equivalency.                       -7.126
## Some.college                                                      -7.401
## Bachelor.s.degree                                                 -6.819
## Master.s.degree                                                   -7.835
## Professional.school.degree                                        -7.771
## Enrolled.In.School                                                -5.582
## Not.Enrolled.In.School                                            -4.629
## In.Armed.Forces                                                   -5.363
## Not.In.labor.force                                                -2.407
## X.10.000.to..14.999                                               -3.720
## X.15.000.to..19.999                                                5.965
## X.25.000.to..29.999                                               -3.981
## X.30.000.to..34.999                                                2.716
## X.40.000.to..44.999                                               -2.799
## X.45.000.to..49.999                                                3.559
## X.50.000.to..59.999                                               -5.149
## X.60.000.to..74.999                                                3.502
## X.75.000.to..99.999                                                5.784
## X.100.000.to..124.999                                             -5.710
## X.125.000.to..149.999                                              4.200
## X.150.000.to..199.999                                              6.143
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.      -4.936
## Average.household.income..In.2014.Inflation.Adjusted.Dollars.      1.948
## Per.capita.income..In.2014.Inflation.adjusted.dollars.             2.256
## Living.in.Poverty                                                  3.921
## At.or.Above.Poverty.Level                                          5.897
```

```
## Crime                                                          -14.209
##                                                               Pr(>|t|)
## (Intercept)                                                    < 2e-16 ***
## Male                                                          0.001173 **
## Under.5.Years                                                 3.37e-05 ***
## X10.to.14.Years                                               0.007348 **
## X15.to.17.Years                                               1.20e-07 ***
## X25.to.34.Years                                               3.61e-06 ***
## X35.to.44.Years                                               1.14e-08 ***
## X45.to.54.Years                                               0.000321 ***
## X55.to.64.Years                                               0.000593 ***
## X65.to.74.Years                                               2.11e-12 ***
## X75.to.84.Years                                               5.07e-15 ***
## White.Alone                                                   7.13e-09 ***
## Black.or.African.American.Alone                               2.10e-09 ***
## Asian.Alone                                                   1.16e-08 ***
## Some.Other.Race.Alone                                         5.39e-09 ***
## Average.Household.Size                                        0.082981 .
## Less.Than.High.School                                         2.71e-13 ***
## High.School.Graduate..includes.equivalency.                   1.16e-12 ***
## Some.college                                                  1.54e-13 ***
## Bachelor.s.degree                                             1.00e-11 ***
## Master.s.degree                                               5.50e-15 ***
## Professional.school.degree                                    9.12e-15 ***
## Enrolled.In.School                                            2.48e-08 ***
## Not.Enrolled.In.School                                        3.76e-06 ***
## In.Armed.Forces                                               8.50e-08 ***
## Not.In.labor.force                                            0.016112 *
## X.10.000.to..14.999                                           0.000201 ***
## X.15.000.to..19.999                                           2.59e-09 ***
## X.25.000.to..29.999                                           6.95e-05 ***
## X.30.000.to..34.999                                           0.006632 **
## X.40.000.to..44.999                                           0.005144 **
## X.45.000.to..49.999                                           0.000375 ***
## X.50.000.to..59.999                                           2.70e-07 ***
## X.60.000.to..74.999                                           0.000465 ***
## X.75.000.to..99.999                                           7.65e-09 ***
## X.100.000.to..124.999                                         1.18e-08 ***
## X.125.000.to..149.999                                         2.71e-05 ***
## X.150.000.to..199.999                                         8.60e-10 ***
## Median.household.income..In.2014.Inflation.Adjusted.Dollars.  8.17e-07 ***
## Average.household.income..In.2014.Inflation.Adjusted.Dollars. 0.051420 .
## Per.capita.income..In.2014.Inflation.adjusted.dollars.        0.024084 *
## Living.in.Poverty                                             8.91e-05 ***
## At.or.Above.Poverty.Level                                     3.90e-09 ***
## Crime                                                          < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4621 on 5956 degrees of freedom
## Multiple R-squared:  0.0637, Adjusted R-squared:  0.05694
## F-statistic: 9.424 on 43 and 5956 DF,  p-value: < 2.2e-16
```

```
# 43 predictors selected

#Calculate R-square of test model
cor(all.train$Duration, predict(lm.step))^2
```

```
## [1] 0.05385125
```

```
# holdout Validation
lm.step.pred<-predict(lm.step, newdata = all.hold)

#Calculate R-square of holdout
cor(all.hold$Duration, lm.step.pred)^2
```

```
## [1] 0.02912706
```

```
# calculate MSE
mean((all.hold$Duration-lm.step.pred)^2)
```

```
## [1] 1603234
```

Cluster-wise regression

```
## [1] 0.05466241
```

```
## [1] 0.02884553
```

Tree models

```
library(rpart)
library(rattle)
```

```
## Rattle: A free graphical interface for data mining with R.
## Version 4.1.0 Copyright (c) 2006-2015 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
#Tree Model on train
set.seed(123)
x.rpart <- rpart(lm.step,data=all.train, control = rpart.control(cp=0, minsplit = 30, xval = 10, maxsur

rpart.pred<- predict(x.rpart, newdata = all.hold)

#Calculate R-square of training set
cor(all.train$Duration, predict(x.rpart))^2
```

```
## [1] 0.1546226
```

```r
#Calculate R-square of holdout
cor(all.hold$Duration, rpart.pred)^2
```

```
## [1] 0.1037972
```

```r
min.xerror <- which.min(x.rpart$cptable[,"xerror"])
min.cp <- x.rpart$cptable[min.xerror,"CP"]

#pruned tree method
z.pruned <- rpart(Duration~.,data=all.train, control = rpart.control(cp=min.cp, minsplit = 10, xval = 1

prune.pred<- predict(z.pruned, newdata = all.hold)

#Calculate R-square of training set
cor(all.train$Duration, predict(z.pruned))^2
```

```
## [1] 0.1702408
```

```r
#Calculate R-square of holdout
cor(all.hold$Duration, prune.pred)^2
```

```
## [1] 0.1190734
```

Random Forest

```r
library(randomForest)
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```r
set.seed(123)
forest <- randomForest(Duration~., data=all.train,importance=TRUE,ntree=20)
for.pred <- predict(forest, all.hold, type="response")

#Calculate R-square of training set
cor(all.train$Duration, predict(forest, type = "response"))^2
```

```
## [1] NA
```

```r
#Calculate R-square of holdout
cor(all.hold$Duration, for.pred)^2
```

```
## [1] 0.1053377
```