Home Mortgage Disclosure Act Data

- **Objective of this Data Analysis: -** HMDA Data provides information regarding home mortgage lending activity. The purpose of analysis the data is that:-

  Will The mortgage application will be accepted or denied?

  The data frame contains 2380 observations on 14 variables.

Deny - Was the mortgage denied
Pirat - Payments to income ratio.
Hirat - Housing expense to income ratio.
Lvrat - Loan to value ratio.
Chist - Credit history: consumer payments.
Mhist - Credit history: mortgage payments.
Phist - Public bad credit record?
Unemp - unemployment rate in applicant's industry
Selfemp- Is the individual self-employed?
Insurance - Was the individual denied mortgage insurance?
Condomin - Is the unit a condominium?
Afam - Is the individual African-American?
Single - Is the individual single?
Hschool - Does the individual have a high-school diploma?

Software and Packaged Used :-

Software :- R Studio

Packages : AER, GGPLOT2, gmodels, Hmisc, caTools, ROCR

Steps Implemented:-

- *Downloading the package*
- *Cleaning the data*
- *Feature Selection*
- *Building Models*
- *Model Selection*
- *Prediction and accuracy*

Code

Description: Cross-section data on the Home Mortgage Disclosure Act (HMDA). A data frame containing 2,380 observations on 14 variables.

Hide

```
library(AER)
data(HMDA)
dataset<- data.frame(HMDA)
names(dataset)
```

```
 [1] "deny"      "pirat"     "hirat"     "lvrat"     "chist"
 [6] "mhist"     "phist"     "unemp"     "selfemp"   "insurance"
[11] "condomin"  "afam"      "single"    "hschool"
```

Hide

```
dim(dataset)
```

```
[1] 2380   14
```

Hide

```
str(dataset)
```

```
'data.frame':   2380 obs. of  14 variables:
 $ deny     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ pirat    : num  0.221 0.265 0.372 0.32 0.36 ...
 $ hirat    : num  0.221 0.265 0.248 0.25 0.35 ...
 $ lvrat    : num  0.8 0.922 0.92 0.86 0.6 ...
 $ chist    : Factor w/ 6 levels "1","2","3","4",..: 5 2 1 1 1 1 1 2 2 2 ...
 $ mhist    : Factor w/ 4 levels "1","2","3","4": 2 2 2 2 1 1 2 2 2 1 ...
 $ phist    : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ unemp    : num  3.9 3.2 3.2 4.3 3.2 ...
 $ selfemp  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ insurance: Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ condomin : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 2 1 1 1 ...
 $ afam     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ single   : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 2 1 1 2 ...
 $ hschool  : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
```

Hide

```
summary(dataset)
```

```
  deny          pirat            hirat            lvrat       
 no :2095   Min.   :0.0000   Min.   :0.0000   Min.   :0.0200  
 yes: 285   1st Qu.:0.2800   1st Qu.:0.2140   1st Qu.:0.6527  
            Median :0.3300   Median :0.2600   Median :0.7795  
            Mean   :0.3308   Mean   :0.2553   Mean   :0.7378  
            3rd Qu.:0.3700   3rd Qu.:0.2988   3rd Qu.:0.8685  
            Max.   :3.0000   Max.   :3.0000   Max.   :1.9500  
 chist    mhist    phist          unemp        selfemp   
 1:1353   1: 747   no :2205   Min.   : 1.800   no :2103  
 2: 441   2:1571   yes: 175   1st Qu.: 3.100   yes: 277  
 3: 126   3:  41              Median : 3.200             
 4:  77   4:  21              Mean   : 3.774             
 5: 182                       3rd Qu.: 3.900             
 6: 201                       Max.   :10.600             
 insurance  condomin    afam      single     hschool  
 no :2332   no :1694   no :2041   no :1444   no :  39  
 yes:  48   yes: 686   yes: 339   yes: 936   yes:2341  
```

Hide

```
head(dataset)
```

	deny <fctr>	pirat <dbl>	hirat <dbl>	lvrat <dbl>	chist <fctr>	mhist <fctr>	phist <fctr>	unemp <dbl>	selfemp <fctr>
1	no	0.221	0.221	0.8000000	5	2	no	3.9	no
2	no	0.265	0.265	0.9218750	2	2	no	3.2	no
3	no	0.372	0.248	0.9203980	1	2	no	3.2	no
4	no	0.320	0.250	0.8604651	1	2	no	4.3	no
5	no	0.360	0.350	0.6000000	1	1	no	3.2	no
6	no	0.240	0.170	0.5105263	1	1	no	3.9	no

6 rows | 1-10 of 14 columns

Hide

```
sapply(dataset,function(x) sum(is.na(x)))
```

```
     deny     pirat     hirat     lvrat     chist     mhist 
        0         0         0         0         0         0 
    phist     unemp   selfemp insurance  condomin      afam 
        0         0         0         0         0         0 
   single   hschool 
        0         0 
```

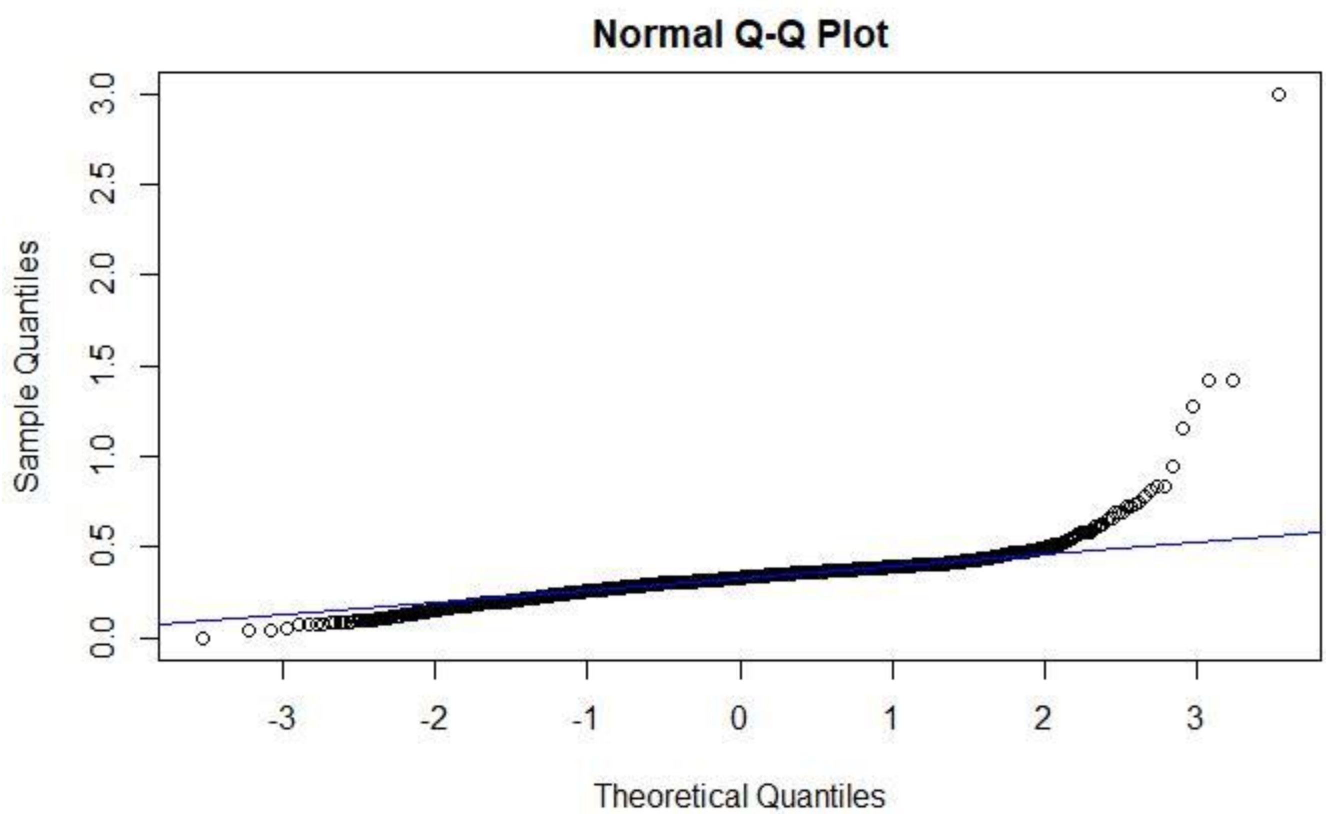
Data is already clean, there are no missing values.

Categorical values are already defined and correctly labeled. pirat,hirat,lvrat,phist are left skewed.

QQ-plot and density plots for payment to income ratio

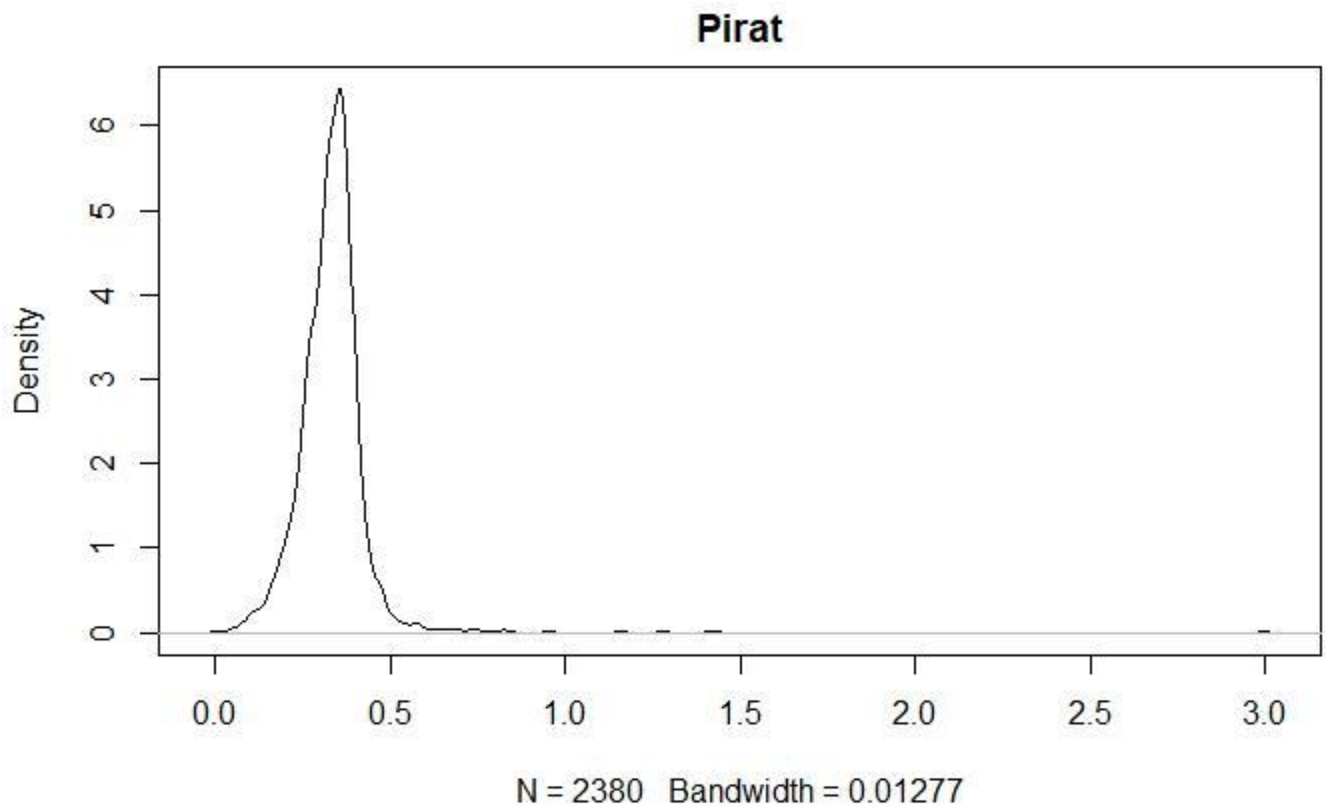
Hide

```
qqnorm(dataset$pirat)
qqline(dataset$pirat,col='blue')
```



Hide

```
plot(density(dataset$pirat),main='Pirat')
```

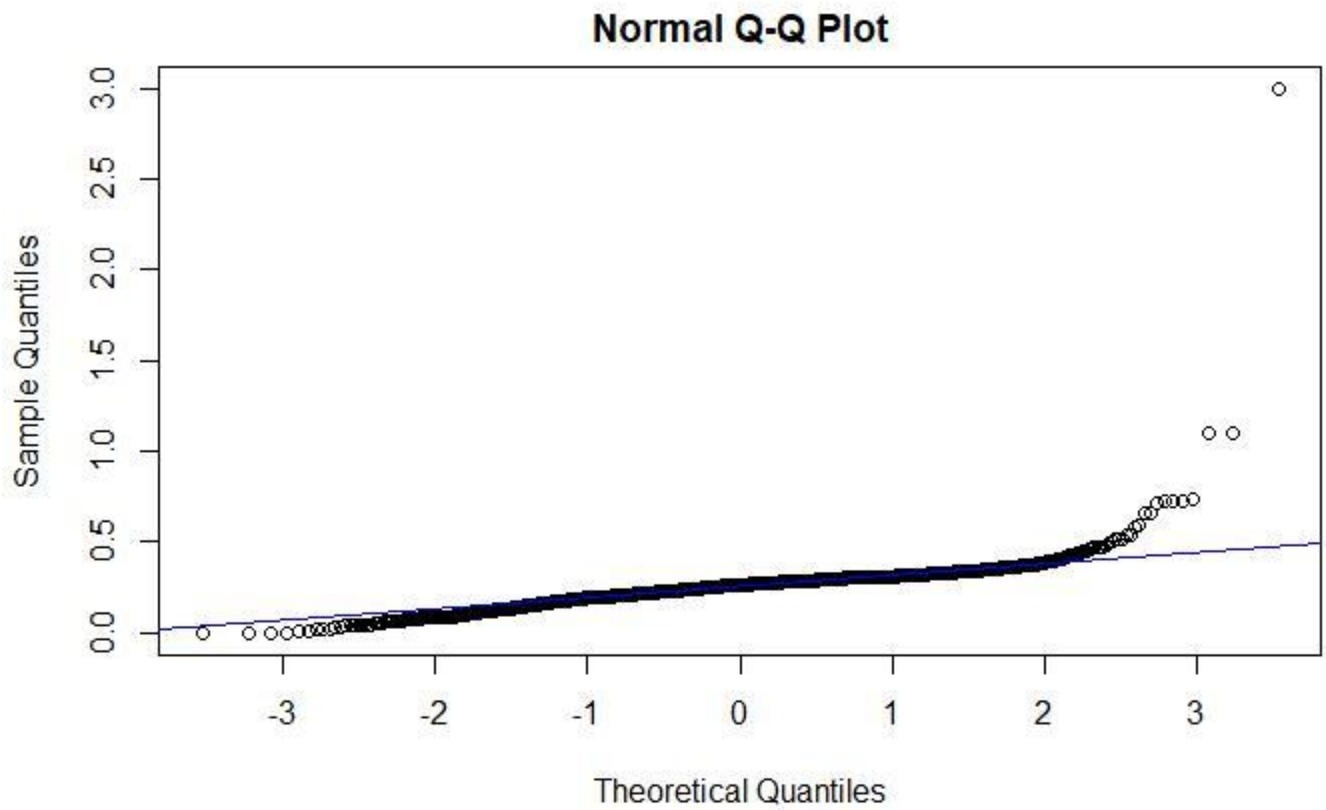


The plot looks right skewed with few outliers.

QQ-plot and density plots housing expense to income ratio

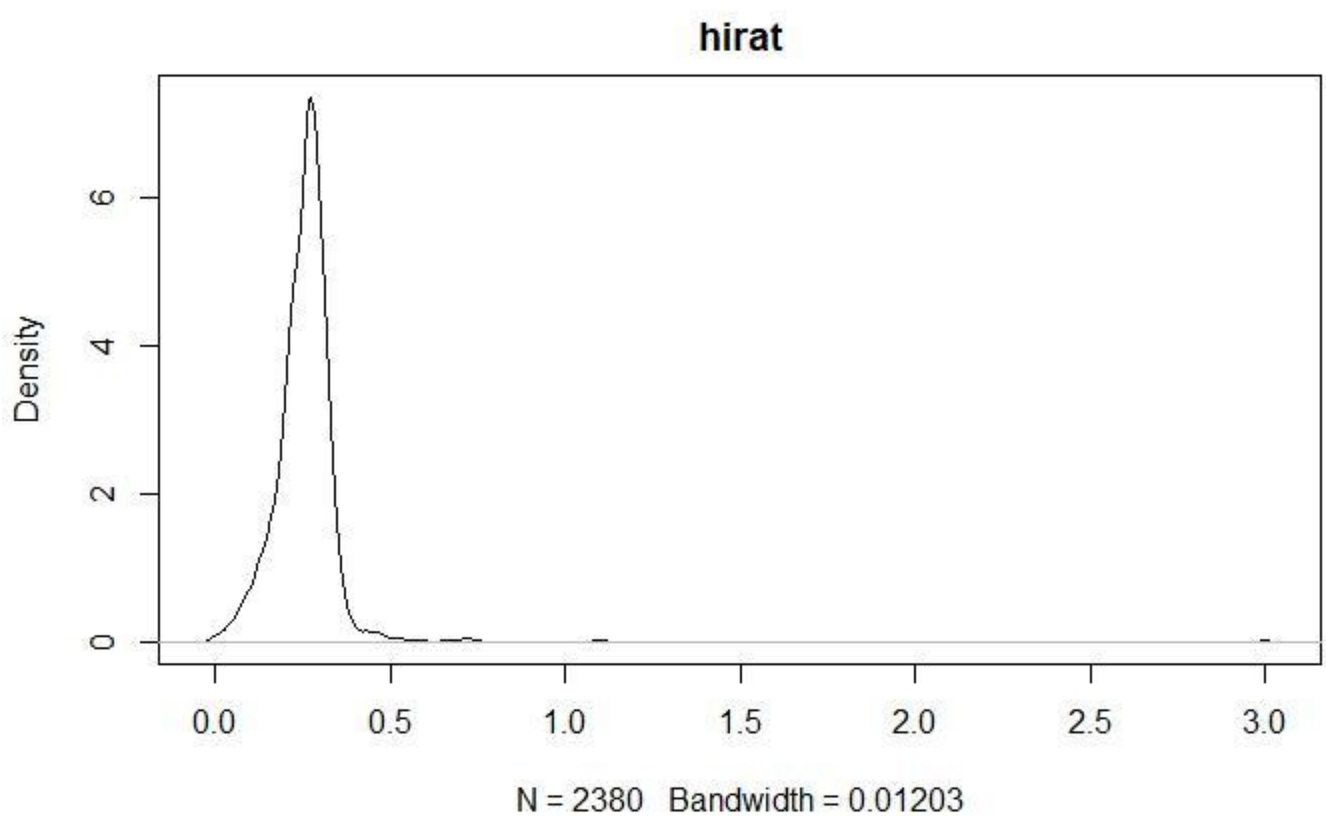
Hide

```
qqnorm(dataset$hirat)
qqline(dataset$hirat,col='blue')
```



Hide

```
plot(density(dataset$hirat),main='hirat')
```

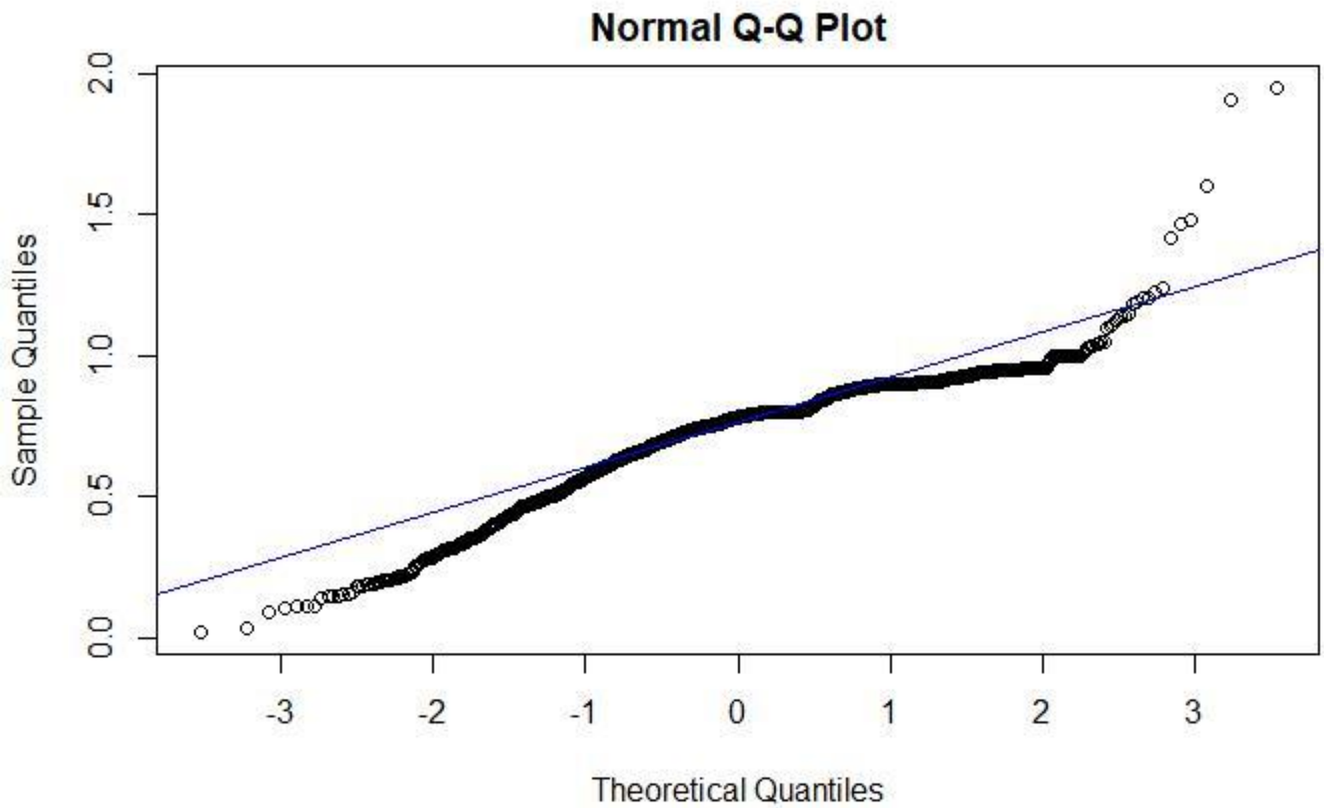


- The plot looks right skewed with few outliers.

- QQ-plot and dnsity plots Loan to value ratio

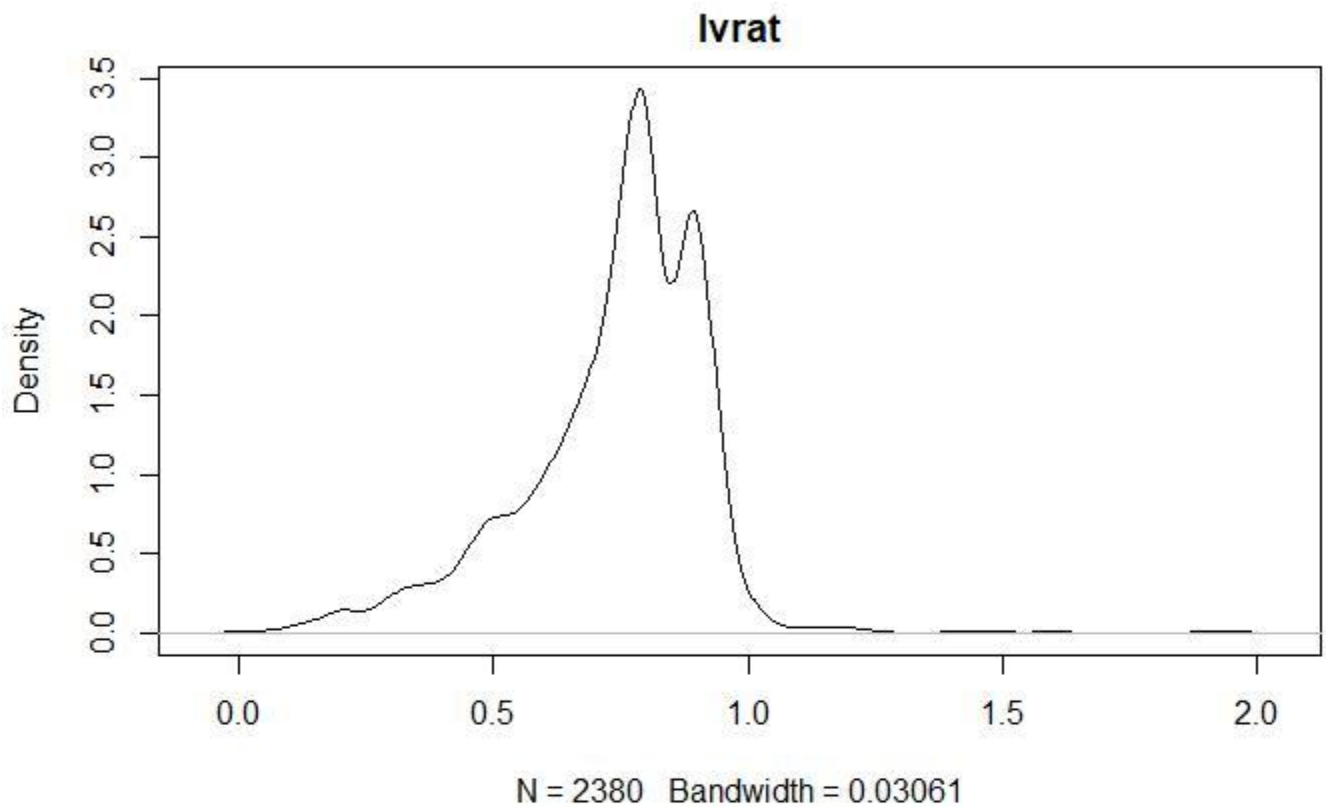
Hide

```
qqnorm(dataset$lvrat)
qqline(dataset$lvrat,col='blue')
```



Hide

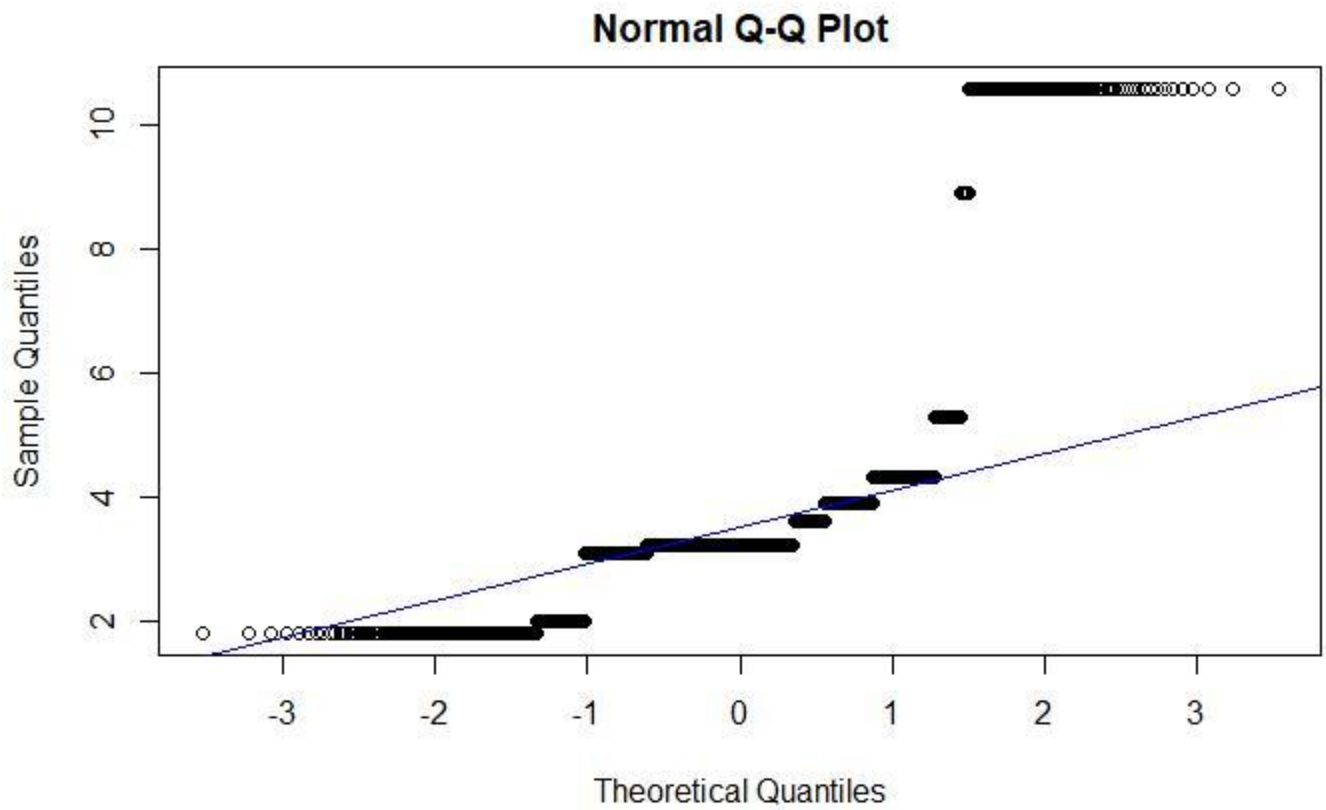
```
plot(density(dataset$lvrat),main='lvrat')
```



QQ-plot and density plots unemployment

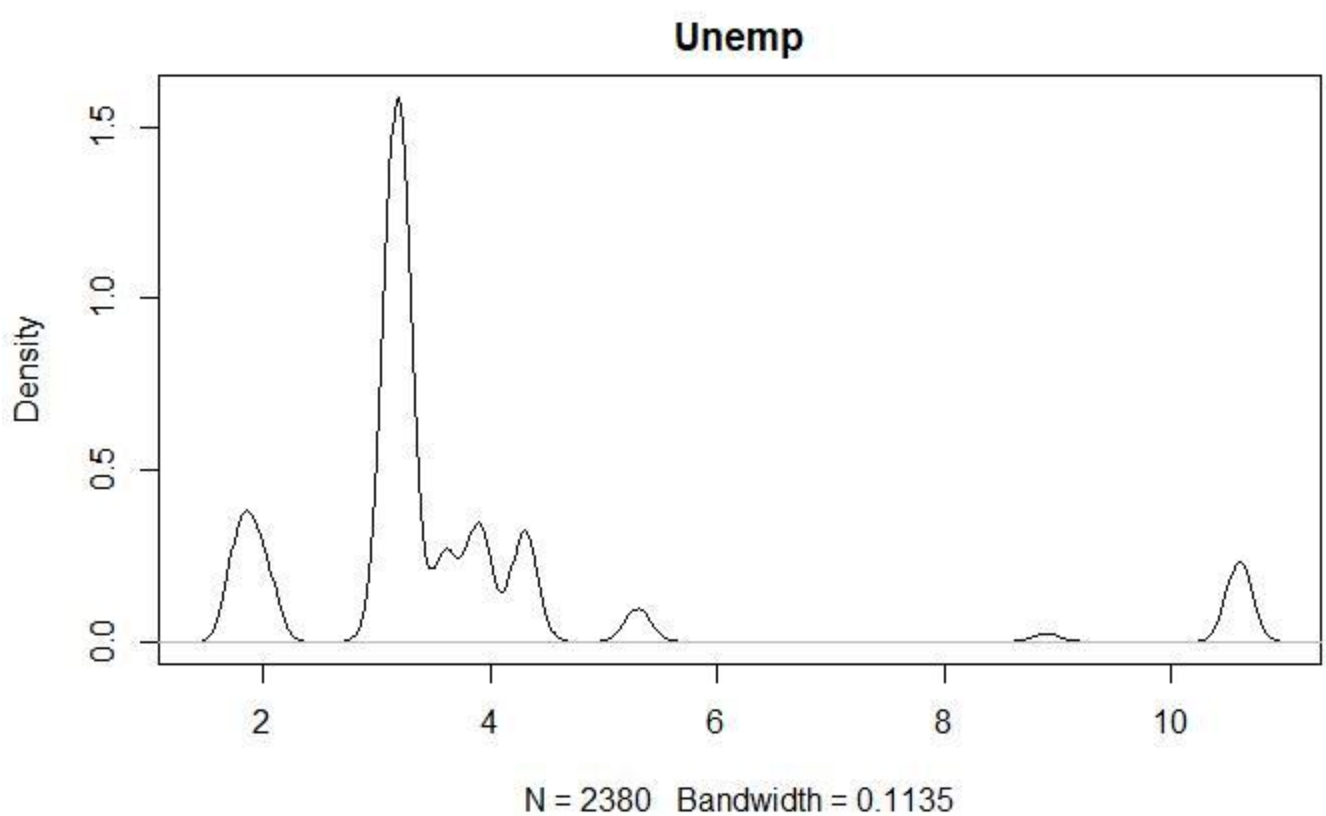
Hide

```
qqnorm(dataset$unemp)
qqline(dataset$unemp,col='blue')
```

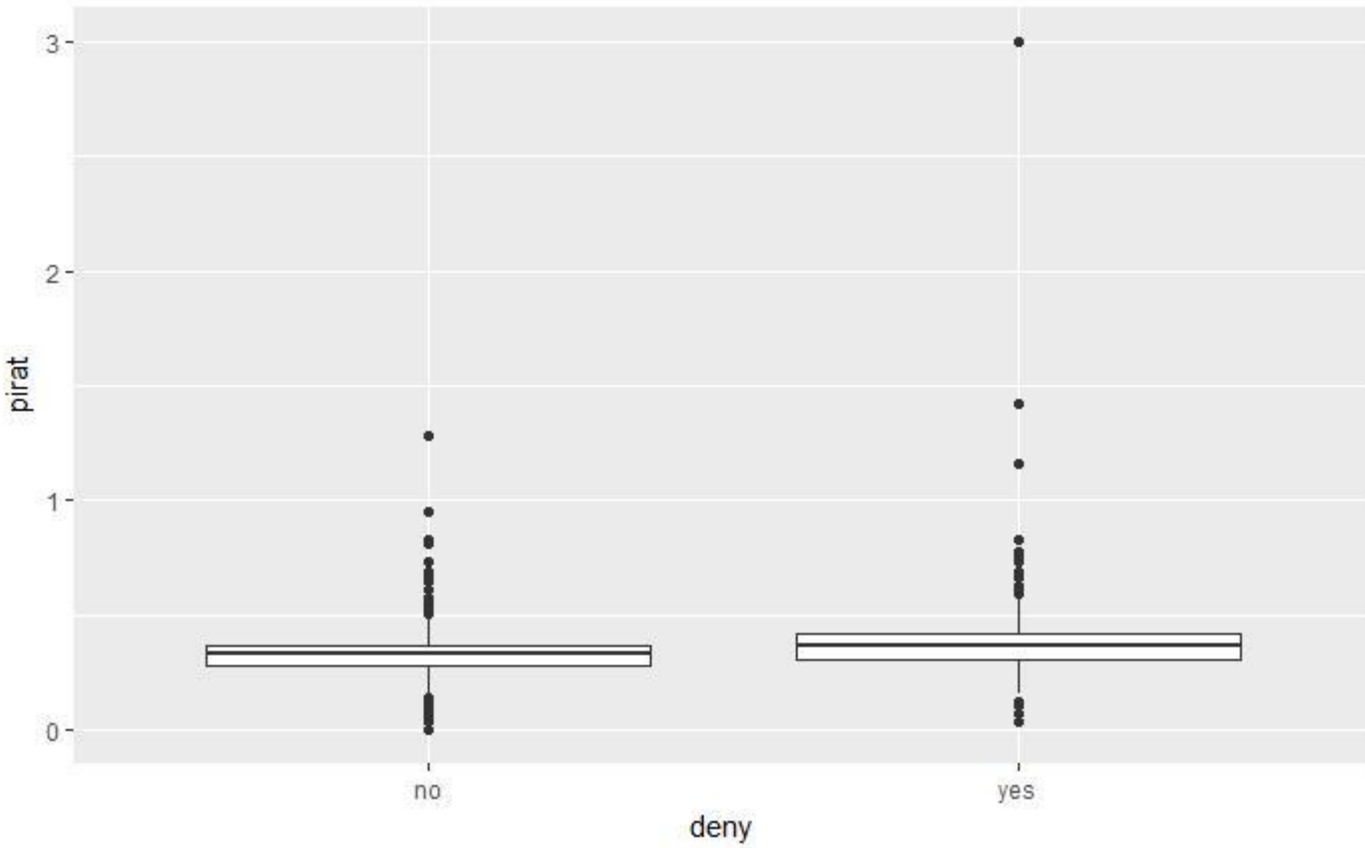


Hide

```
plot(density(dataset$unemp),main='Unemp')
```



Box-Plots



From the various QQ-Plots and Box-Plots we can conclude outliers are present.

Model fitting:

Splitting the data into training and testing set. The training set will be used to fit our model which we will be testing over the testing set.

Hide

```
dt = sort(sample(nrow(dataset), nrow(dataset)*.8))
train<-dataset[dt,]
test<-dataset[-dt,]
```

Now, let’s fit the model.

Hide

```
model1 <- glm(deny~.,family=binomial(link='logit'),data=train)
summary(model1)
```

```
Call:
glm(formula = deny ~ ., family = binomial(link = "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6476  -0.4387  -0.2854  -0.2012   3.0723

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.52897    0.75864  -7.288 3.15e-13 ***
pirat          5.80530    1.30792   4.439 9.06e-06 ***
hirat         -1.08383    1.44386  -0.751  0.45286
lvrat          1.54111    0.56121   2.746  0.00603 **
chist2         0.92487    0.23243   3.979 6.92e-05 ***
chist3         0.94968    0.35320   2.689  0.00717 **
chist4         1.52758    0.39501   3.867  0.00011 ***
chist5         1.30247    0.28064   4.641 3.47e-06 ***
chist6         1.51813    0.26512   5.726 1.03e-08 ***
mhist2         0.30596    0.22043   1.388  0.16512
mhist3         0.36485    0.55327   0.659  0.50960
mhist4         0.99035    0.67127   1.475  0.14012
phistyes       1.39779    0.23370   5.981 2.22e-09 ***
unemp          0.04733    0.03810   1.242  0.21418
selfempyes     0.68878    0.24443   2.818  0.00483 **
insuranceyes   4.38739    0.58180   7.541 4.66e-14 ***
condominyes    0.03347    0.19243   0.174  0.86193
afamyes        0.69662    0.20528   3.393  0.00069 ***
singleyes      0.33779    0.17979   1.879  0.06026 .
hschoolyes    -1.26681    0.47112  -2.689  0.00717 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance:  995.3  on 1884  degrees of freedom
AIC: 1035.3

Number of Fisher Scoring iterations: 6
```

By using function summary() we obtain the results of our model. We are using the Logisitic Regression because the response variable is Categorical(Yes/No) . We can see that hirat, mhist3,mhist4,unemp and condominyes are not statistically significant. Whereas pirat, phistyes, insuranceyes, afamyes are statistically significant variables based on the p-values and AIC is 1024.9.

Now we can run the anova() function on the model to analyze the table of deviance

Hide

```
anova(model1, test="Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: deny

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)    
NULL                       1903     1379.3              
pirat      1   69.560      1902     1309.7 < 2.2e-16 ***
hirat      1    1.006      1901     1308.7 0.3159432    
lvrat      1   39.427      1900     1269.3 3.405e-10 ***
chist      5  102.197      1895     1167.1 < 2.2e-16 ***
mhist      3    4.995      1892     1162.1 0.1721352    
phist      1   43.250      1891     1118.9 4.817e-11 ***
unemp      1    2.349      1890     1116.5 0.1253956    
selfemp    1    5.738      1889     1110.8 0.0166010 *  
insurance  1   92.610      1888     1018.2 < 2.2e-16 ***
condomin   1    1.597      1887     1016.6 0.2063464    
afam       1   12.533      1886     1004.0 0.0003999 ***
single     1    2.362      1885     1001.7 0.1243409    
hschool    1    6.373      1884      995.3 0.0115867 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The difference between the null deviance and the residual deviance shows how our model is doing against the null model (a model with only the intercept). The wider this gap, the better. Analyzing the table we can see the drop in deviance when adding each variable one at a time. Again, adding lvrat, chist,phist,afam and hschool significantly reduces the residual deviance. A large p-value here indicates that the model without the variable explains more or less the same amount of variation. Ultimately what you would like to see is a significant drop in deviance and the AIC.

Hide

```
model2 <- glm(deny~pirat+lvrat+chist+phist+selfemp+insurance+afam+single+hschool,family=binomial
(link='logit'),data=train)
summary(model2)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + selfemp + 
    insurance + afam + single + hschool, family = binomial(link = 
    "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7767  -0.4372  -0.2889  -0.2092   3.1266  

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)    -5.1610     0.7124  -7.245 4.32e-13 ***
pirat           5.0757     0.9797   5.181 2.21e-07 ***
lvrat           1.6481     0.5547   2.971 0.002967 ** 
chist2          0.9274     0.2304   4.024 5.71e-05 ***
chist3          0.9817     0.3464   2.834 0.004590 ** 
chist4          1.5361     0.3907   3.931 8.44e-05 ***
chist5          1.3342     0.2797   4.770 1.84e-06 ***
chist6          1.5994     0.2599   6.153 7.60e-10 ***
phistyes        1.3966     0.2319   6.023 1.72e-09 ***
selfempyes      0.6987     0.2393   2.920 0.003498 ** 
insuranceyes    4.3937     0.5789   7.590 3.21e-14 ***
afamyes         0.6915     0.1999   3.460 0.000541 ***
singleyes       0.3655     0.1722   2.123 0.033784 *  
hschoolyes     -1.3481     0.4714  -2.859 0.004243 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1000.7  on 1890  degrees of freedom
AIC: 1028.7

Number of Fisher Scoring iterations: 6
```

In model two hirat,unemp,mhist and condomin are removed. AIC is 1016.6.

Hide

```
anova(model2, test="Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: deny

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)    
NULL                       1903     1379.3              
pirat      1   69.560      1902     1309.7 < 2.2e-16 ***
lvrat      1   39.031      1901     1270.7 4.172e-10 ***
chist      5  103.087      1896     1167.6 < 2.2e-16 ***
phist      1   42.455      1895     1125.2 7.232e-11 ***
selfemp    1    5.368      1894     1119.8 0.0205144 *  
insurance  1   94.509      1893     1025.3 < 2.2e-16 ***
afam       1   14.200      1892     1011.1 0.0001643 ***
single     1    3.284      1891     1007.8 0.0699650 .  
hschool    1    7.126      1890     1000.7 0.0075956 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
model3 <- glm(deny~pirat+lvrat+chist+phist+insurance+afam+single,family=binomial(link='logit'),d
ata=train)
summary(model3)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance + afam + 
    single, family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7501  -0.4456  -0.2975  -0.2184   3.0892  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.3005     0.5697 -11.060  < 2e-16 ***
pirat          5.1992     0.9756   5.329 9.86e-08 ***
lvrat          1.5189     0.5573   2.725 0.006425 ** 
chist2         0.9398     0.2291   4.103 4.08e-05 ***
chist3         0.9981     0.3404   2.932 0.003368 ** 
chist4         1.5950     0.3865   4.127 3.68e-05 ***
chist5         1.3707     0.2772   4.944 7.63e-07 ***
chist6         1.5318     0.2577   5.944 2.78e-09 ***
phistyes       1.4283     0.2304   6.199 5.68e-10 ***
insuranceyes   4.3203     0.5743   7.522 5.39e-14 ***
afamyes        0.6933     0.1967   3.525 0.000423 ***
singleyes      0.3038     0.1692   1.796 0.072552 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1015.2  on 1892  degrees of freedom
AIC: 1039.2

Number of Fisher Scoring iterations: 6
```

Hide

```
anova(model3, test="Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: deny

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)    
NULL                       1903     1379.3              
pirat      1   69.560      1902     1309.7 < 2.2e-16 ***
lvrat      1   39.031      1901     1270.7 4.172e-10 ***
chist      5  103.087      1896     1167.6 < 2.2e-16 ***
phist      1   42.455      1895     1125.2 7.232e-11 ***
insurance  1   93.852      1894     1031.3 < 2.2e-16 ***
afam       1   12.923      1893     1018.4 0.0003246 ***
single     1    3.208      1892     1015.2 0.0732927 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Even we can see the p-value associated with selfemp and unemp is not significant as its large but removing the element from model increases the AIC value. So model3 is not a good model.

From here we conclude that model2 is the best. Now we will use forward selection to verify our model.

Hide

```
fit1<- glm(deny~pirat,family=binomial(link='logit'),data=train)
summary(fit1)
```

```
Call:
glm(formula = deny ~ pirat, family = binomial(link =
    "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9766  -0.5231  -0.4645  -0.3761   2.7466

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.1900     0.3138 -13.351  < 2e-16 ***
pirat         6.3059     0.8593   7.339 2.16e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1309.7 on 1902  degrees of freedom
AIC: 1313.7

Number of Fisher Scoring iterations: 5
```

Hide

```
fit2<-glm(deny~pirat+hirat,family=binomial(link='logit'),data=train)
summary(fit2)
```

```
Call:
glm(formula = deny ~ pirat + hirat, family = binomial(link = 
    "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.0596  -0.5267  -0.4605   -0.3794   2.7090  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -4.1078     0.3224 -12.743  < 2e-16 ***
pirat         6.9753     1.0829   6.441 1.18e-10 ***
hirat        -1.1943     1.1831  -1.010    0.313    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1308.7 on 1901  degrees of freedom
AIC: 1314.7

Number of Fisher Scoring iterations: 5
```

Hide

```
fit3<-glm(deny~pirat+lvrat,family=binomial(link='logit'),data=train)
summary(fit3)
```

```

Call:
glm(formula = deny ~ pirat + lvrat, family = binomial(link =
    "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7143  -0.5443  -0.4418   -0.3050   3.0726

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.3406     0.5091 -12.456  < 2e-16 ***
pirat         6.0533     0.8802   6.877 6.10e-12 ***
lvrat         2.8771     0.4868   5.911 3.41e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1270.7 on 1901  degrees of freedom
AIC: 1276.7

Number of Fisher Scoring iterations: 5
```

Hide

```
fit4<-glm(deny~pirat+lvrat+chist,family=binomial(link='logit'),data=train)
summary(fit4)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist, family = binomial(link =
    "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1041  -0.4913  -0.3524   -0.2403   3.2687

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.8765     0.5410 -12.710  < 2e-16 ***
pirat         5.7497     0.9369   6.137 8.40e-10 ***
lvrat         2.6879     0.5014   5.361 8.29e-08 ***
chist2        0.9233     0.2102   4.393 1.12e-05 ***
chist3        1.2589     0.3027   4.158 3.21e-05 ***
chist4        1.5415     0.3722   4.142 3.45e-05 ***
chist5        1.4543     0.2516   5.780 7.48e-09 ***
chist6        2.0417     0.2202   9.270  < 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: 1379.3 on 1903  degrees of freedom
Residual deviance: 1167.6 on 1896  degrees of freedom
AIC: 1183.6

Number of Fisher Scoring iterations: 6
```

Hide

```
fit5<-glm(deny~pirat+lvrat+chist+mhist,family=binomial(link='logit'),data=train)
summary(fit5)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + mhist, family = binomial(link = 
    "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.1323  -0.4933  -0.3537  -0.2322   3.2154  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -7.0499     0.5528 -12.753  < 2e-16 ***
pirat         5.8429     0.9452   6.181 6.35e-10 ***
lvrat         2.5290     0.5089   4.969 6.72e-07 ***
chist2        0.9424     0.2109   4.469 7.87e-06 ***
chist3        1.2593     0.3077   4.093 4.25e-05 ***
chist4        1.4895     0.3734   3.989 6.63e-05 ***
chist5        1.4351     0.2521   5.692 1.26e-08 ***
chist6        1.9728     0.2234   8.832  < 2e-16 ***
mhist2        0.3602     0.1927   1.869   0.0616 .  
mhist3        0.3074     0.5366   0.573   0.5667    
mhist4        0.8871     0.6147   1.443   0.1490    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1163.0  on 1893  degrees of freedom
AIC: 1185

Number of Fisher Scoring iterations: 6
```

Hide

```
fit6<-glm(deny~pirat+lvrat+chist+phist,family=binomial(link='logit'),data=train)
summary(fit6)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist, family = binomial(link =
    "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8198  -0.4831  -0.3406   -0.2348   3.2576

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.7726     0.5469 -12.383  < 2e-16 ***
pirat         5.5239     0.9433   5.856 4.75e-09 ***
lvrat         2.5452     0.5061   5.029 4.94e-07 ***
chist2        0.8874     0.2127   4.173 3.01e-05 ***
chist3        1.0696     0.3106   3.444 0.000573 ***
chist4        1.5846     0.3747   4.229 2.35e-05 ***
chist5        1.2978     0.2590   5.011 5.43e-07 ***
chist6        1.5766     0.2395   6.582 4.63e-11 ***
phistyes      1.4599     0.2173   6.719 1.84e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1125.2 on 1895  degrees of freedom
AIC: 1143.2

Number of Fisher Scoring iterations: 6
```

Hide

```
fit7<-glm(deny~pirat+lvrat+chist+phist+unemp,family=binomial(link='logit'),data=train)
summary(fit7)
```

```

Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + unemp, family = binomial(link =
    "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9799  -0.4792  -0.3388  -0.2304   3.2924

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.99885    0.56712 -12.341  < 2e-16 ***
pirat        5.49131    0.94623   5.803 6.50e-09 ***
lvrat        2.57236    0.50583   5.085 3.67e-07 ***
chist2       0.87272    0.21299   4.097 4.18e-05 ***
chist3       1.05813    0.31076   3.405 0.000662 ***
chist4       1.60088    0.37514   4.267 1.98e-05 ***
chist5       1.29132    0.25970   4.972 6.61e-07 ***
chist6       1.58660    0.23987   6.614 3.73e-11 ***
phistyes     1.44906    0.21728   6.669 2.57e-11 ***
unemp        0.05627    0.03465   1.624 0.104349
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1122.7 on 1894  degrees of freedom
AIC: 1142.7

Number of Fisher Scoring iterations: 6
```

Hide

```
fit8<-glm(deny~pirat+lvrat+chist+phist+selfemp,family=binomial(link='logit'),data=train)
summary(fit8)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + selfemp,
    family = binomial(link = "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0204  -0.4775  -0.3344   -0.2314   3.2893

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.8783     0.5471 -12.571  < 2e-16 ***
pirat         5.4512     0.9402   5.798 6.71e-09 ***
lvrat         2.6232     0.5061   5.183 2.19e-07 ***
chist2        0.8785     0.2131   4.123 3.74e-05 ***
chist3        1.0579     0.3116   3.395 0.000686 ***
chist4        1.5807     0.3771   4.192 2.77e-05 ***
chist5        1.2962     0.2599   4.986 6.15e-07 ***
chist6        1.6218     0.2406   6.740 1.58e-11 ***
phistyes      1.4384     0.2173   6.621 3.58e-11 ***
selfempyes    0.5447     0.2275   2.395 0.016642 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1119.8  on 1894  degrees of freedom
AIC: 1139.8

Number of Fisher Scoring iterations: 6
```

Hide

```
fit9<-glm(deny~pirat+lvrat+chist+phist+insurance,family=binomial(link='logit'),data=train)
summary(fit9)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance, 
    family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.6882  -0.4496  -0.3074  -0.2229   3.1935  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.4220     0.5676 -11.315  < 2e-16 ***
pirat          5.4612     0.9736   5.609 2.03e-08 ***
lvrat          1.8440     0.5487   3.361 0.000778 ***
chist2         0.9468     0.2278   4.156 3.24e-05 ***
chist3         1.0677     0.3380   3.159 0.001585 ** 
chist4         1.8046     0.3784   4.769 1.85e-06 ***
chist5         1.4437     0.2711   5.326 1.01e-07 ***
chist6         1.7070     0.2520   6.773 1.26e-11 ***
phistyes       1.4757     0.2264   6.519 7.08e-11 ***
insuranceyes   4.2638     0.5689   7.495 6.61e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1031.3  on 1894  degrees of freedom
AIC: 1051.3

Number of Fisher Scoring iterations: 6
```

Hide

```
anova(fit9, test="Chisq")
```

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: deny

Terms added sequentially (first to last)

          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                       1903     1379.3
pirat      1   69.560      1902     1309.7< 2.2e-16 ***
lvrat      1   39.031      1901     1270.74.172e-10 ***
chist      5  103.087      1896     1167.6 < 2.2e-16***
phist      1   42.455      1895     1125.2 7.232e-11***
insurance  1   93.852      1894     1031.3 < 2.2e-16***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
fit10<-glm(deny~pirat+lvrat+chist+phist+insurance+condomin,family=binomial(link='logit'),data=tr
ain)
summary(fit10)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance + 
    condomin, family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7330  -0.4509  -0.3056  -0.2218   3.1360  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.4555     0.5674 -11.378  < 2e-16 ***
pirat          5.4794     0.9740   5.626 1.85e-08 ***
lvrat          1.7893     0.5494   3.257  0.00113 ** 
chist2         0.9671     0.2286   4.231 2.32e-05 ***
chist3         1.0716     0.3379   3.171  0.00152 ** 
chist4         1.7517     0.3811   4.596 4.31e-06 ***
chist5         1.4407     0.2712   5.312 1.09e-07 ***
chist6         1.6867     0.2529   6.670 2.56e-11 ***
phistyes       1.4778     0.2270   6.509 7.56e-11 ***
insuranceyes   4.2807     0.5690   7.523 5.37e-14 ***
condominyes    0.2238     0.1787   1.252  0.21052    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1029.8  on 1893  degrees of freedom
AIC: 1051.8

Number of Fisher Scoring iterations: 6
```

Hide

```
fit11<-glm(deny~pirat+lvrat+chist+phist+insurance+afam,family=binomial(link='logit'),data=train)
summary(fit11)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance
    + afam, family = binomial(link = "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8061  -0.4461  -0.2995  -0.2207   3.0578

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.2389     0.5689 -10.966  < 2e-16 ***
pirat          5.3122     0.9759   5.443 5.23e-08 ***
lvrat          1.5546     0.5586   2.783 0.005386 **
chist2         0.9279     0.2285   4.060 4.90e-05 ***
chist3         0.9964     0.3389   2.940 0.003284 **
chist4         1.6406     0.3851   4.260 2.05e-05 ***
chist5         1.3548     0.2761   4.906 9.28e-07 ***
chist6         1.5417     0.2579   5.979 2.25e-09 ***
phistyes       1.4117     0.2300   6.138 8.37e-10 ***
insuranceyes   4.3049     0.5719   7.528 5.17e-14 ***
afamyes        0.7234     0.1961   3.690 0.000224 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1018.4  on 1893  degrees of freedom
AIC: 1040.4

Number of Fisher Scoring iterations: 6
```

Hide

```
fit12<-glm(deny~pirat+lvrat+chist+phist+insurance+afam+single,family=binomial(link='logit'),data
=train)
summary(fit12)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance + afam + 
    single, family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7501  -0.4456  -0.2975  -0.2184   3.0892  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.3005    0.5697 -11.060  < 2e-16 ***
pirat          5.1992    0.9756   5.329 9.86e-08 ***
lvrat          1.5189    0.5573   2.725 0.006425 ** 
chist2         0.9398    0.2291   4.103 4.08e-05 ***
chist3         0.9981    0.3404   2.932 0.003368 ** 
chist4         1.5950    0.3865   4.127 3.68e-05 ***
chist5         1.3707    0.2772   4.944 7.63e-07 ***
chist6         1.5318    0.2577   5.944 2.78e-09 ***
phistyes       1.4283    0.2304   6.199 5.68e-10 ***
insuranceyes   4.3203    0.5743   7.522 5.39e-14 ***
afamyes        0.6933    0.1967   3.525 0.000423 ***
singleyes      0.3038    0.1692   1.796 0.072552 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1015.2  on 1892  degrees of freedom
AIC: 1039.2

Number of Fisher Scoring iterations: 6
```

Hide

```
fit13<-glm(deny~pirat+lvrat+chist+insurance+afam+single+hschool,family=binomial(link='logit'),da
ta=train)
summary(fit13)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + insurance + afam + single + 
    hschool, family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q       Max  
-2.8563  -0.4638  -0.3053  -0.2199    3.0927  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -5.1776     0.7036 -7.358 1.86e-13 ***
pirat          5.3940     0.9632  5.600 2.14e-08 ***
lvrat          1.6977     0.5511  3.081 0.002065 ** 
chist2         0.9773     0.2274  4.298 1.72e-05 ***
chist3         1.1881     0.3350  3.547 0.000390 ***
chist4         1.4877     0.3852  3.862 0.000112 ***
chist5         1.4895     0.2697  5.523 3.33e-08 ***
chist6         1.9685     0.2411  8.163 3.26e-16 ***
insuranceyes   4.3051     0.5715  7.533 4.97e-14 ***
afamyes        0.7462     0.1917  3.892 9.94e-05 ***
singleyes      0.3127     0.1679  1.863 0.062477 .  
hschoolyes    -1.2859     0.4644 -2.769 0.005627 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1044.4  on 1892  degrees of freedom
AIC: 1068.4

Number of Fisher Scoring iterations: 6
```

fit2 discard as AIC increases and p-value is too large fit3 is good AIC reduces fit4
fit5 even though the p-values are too large the AIC of the model decreases** fit6 is good fit7 although the AIC remains same we can see thr p-value associated with unemp is large so fit7 is not a good model fit8 AIC remains same, the p-value is greater than 0.05 so fit8 is discarded fit9 Even though the AIC increases, but the deviance decreases significantly and the assosiated p-value is significant we will keep fit9 fit10 discard based on p-values fit11 is good reduces AIC a lot fit12 is good fit13 discard

So fit12 is the best model. Comparing it with our previous model3

Hide

```
model3 <- glm(deny~pirat+lvrat+chist+phist+insurance+afam+single,family=binomial(link='logit'),d
ata=train)
summary(model3)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance + afam +
    single, family = binomial(link = "logit"), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7501  -0.4456  -0.2975  -0.2184   3.0892

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -6.3005     0.5697 -11.060  < 2e-16 ***
pirat           5.1992     0.9756   5.329 9.86e-08 ***
lvrat           1.5189     0.5573   2.725 0.006425 **
chist2          0.9398     0.2291   4.103 4.08e-05 ***
chist3          0.9981     0.3404   2.932 0.003368 **
chist4          1.5950     0.3865   4.127 3.68e-05 ***
chist5          1.3707     0.2772   4.944 7.63e-07 ***
chist6          1.5318     0.2577   5.944 2.78e-09 ***
phistyes        1.4283     0.2304   6.199 5.68e-10 ***
insuranceyes    4.3203     0.5743   7.522 5.39e-14 ***
afamyes         0.6933     0.1967   3.525 0.000423 ***
singleyes       0.3038     0.1692   1.796 0.072552 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3 on 1903  degrees of freedom
Residual deviance: 1015.2 on 1892  degrees of freedom
AIC: 1039.2

Number of Fisher Scoring iterations: 6
```

Hide

```
fit12<-glm(deny~pirat+lvrat+chist+phist+insurance+afam+single,family=binomial(link='logit'),data
=train)
summary(fit12)
```

```
Call:
glm(formula = deny ~ pirat + lvrat + chist + phist + insurance + afam + 
    single, family = binomial(link = "logit"), data = train)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7501  -0.4456  -0.2975  -0.2184   3.0892  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.3005     0.5697 -11.060  < 2e-16 ***
pirat          5.1992     0.9756   5.329 9.86e-08 ***
lvrat          1.5189     0.5573   2.725 0.006425 ** 
chist2         0.9398     0.2291   4.103 4.08e-05 ***
chist3         0.9981     0.3404   2.932 0.003368 ** 
chist4         1.5950     0.3865   4.127 3.68e-05 ***
chist5         1.3707     0.2772   4.944 7.63e-07 ***
chist6         1.5318     0.2577   5.944 2.78e-09 ***
phistyes       1.4283     0.2304   6.199 5.68e-10 ***
insuranceyes   4.3203     0.5743   7.522 5.39e-14 ***
afamyes        0.6933     0.1967   3.525 0.000423 ***
singleyes      0.3038     0.1692   1.796 0.072552 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1379.3  on 1903  degrees of freedom
Residual deviance: 1015.2  on 1892  degrees of freedom
AIC: 1039.2

Number of Fisher Scoring iterations: 6
```

So from the forward model selection method we can conclude Model3 is the best.

So our model is : model3 <- glm(deny~pirat+lvrat+chist+phist+insurance+afam+single,family=binomial(link='logit'),data=train)

Hide

```
fm1 <- lm(I(as.numeric(deny) - 1) ~ pirat+lvrat+chist+phist+insurance+afam+single, data = datase
t)
summary(fm1)
```

```
Call:
lm(formula = I(as.numeric(deny) - 1) ~ pirat + lvrat + chist
    + phist + insurance + afam + single, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-0.93240 -0.11506 -0.05449 -0.00940  1.08223

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.18538    0.02903  -6.386 2.04e-10 ***
pirat         0.43341    0.05461   7.937 3.16e-15 ***
lvrat         0.09423    0.03327   2.832  0.00466 **
chist2        0.04236    0.01541   2.750  0.00601 **
chist3        0.05243    0.02643   1.984  0.04738 *
chist4        0.13951    0.03316   4.208 2.67e-05 ***
chist5        0.11345    0.02246   5.051 4.74e-07 ***
chist6        0.16169    0.02255   7.170 9.95e-13 ***
phistyes      0.20813    0.02357   8.829  < 2e-16 ***
insuranceyes  0.71292    0.04143  17.207  < 2e-16 ***
afamyes       0.08340    0.01728   4.826 1.48e-06 ***
singleyes     0.03355    0.01188   2.823  0.00479 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2806 on 2368 degrees of freedom
Multiple R-squared:  0.2567,    Adjusted R-squared:  0.2533
F-statistic: 74.36 on 11 and 2368 DF,  p-value: < 2.2e-16
```

Hide

```
fm2 <- lm(I(as.numeric(deny) - 1) ~ pirat+lvrat+chist+insurance+afam+single+hschool, data = data
set)
summary(fm2)
```

```

Call:
lm(formula = I(as.numeric(deny) - 1) ~ pirat + lvrat + chist
    + insurance + afam + single + hschool, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01018 -0.12217 -0.05868 -0.00999  1.08383

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.07432    0.05553  -1.338  0.18096
pirat          0.45315    0.05552   8.163 5.27e-16 ***
lvrat          0.10746    0.03373   3.186  0.00146 **
chist2         0.04638    0.01563   2.968  0.00303 **
chist3         0.07994    0.02664   3.001  0.00272 **
chist4         0.13161    0.03369   3.906 9.64e-05 ***
chist5         0.13060    0.02270   5.753 9.87e-09 ***
chist6         0.21777    0.02200   9.900  < 2e-16 ***
insuranceyes   0.73613    0.04199  17.531  < 2e-16 ***
afamyes        0.09489    0.01749   5.427 6.33e-08 ***
singleyes      0.03379    0.01208   2.797  0.00519 **
hschoolyes    -0.12428    0.04641  -2.678  0.00746 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2848 on 2368 degrees of freedom
Multiple R-squared:  0.2346,    Adjusted R-squared:  0.231
F-statistic: 65.98 on 11 and 2368 DF,  p-value: < 2.2e-16
```

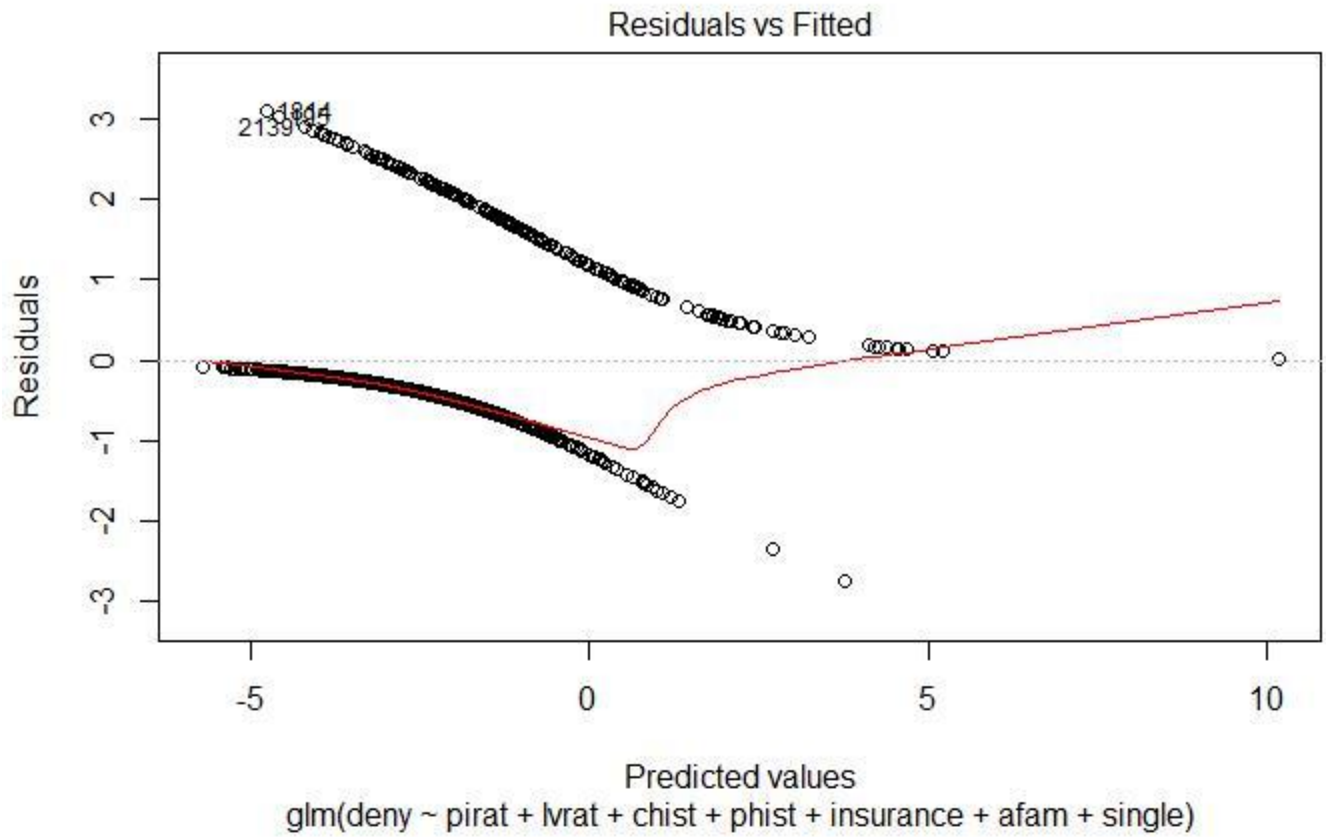
Hide

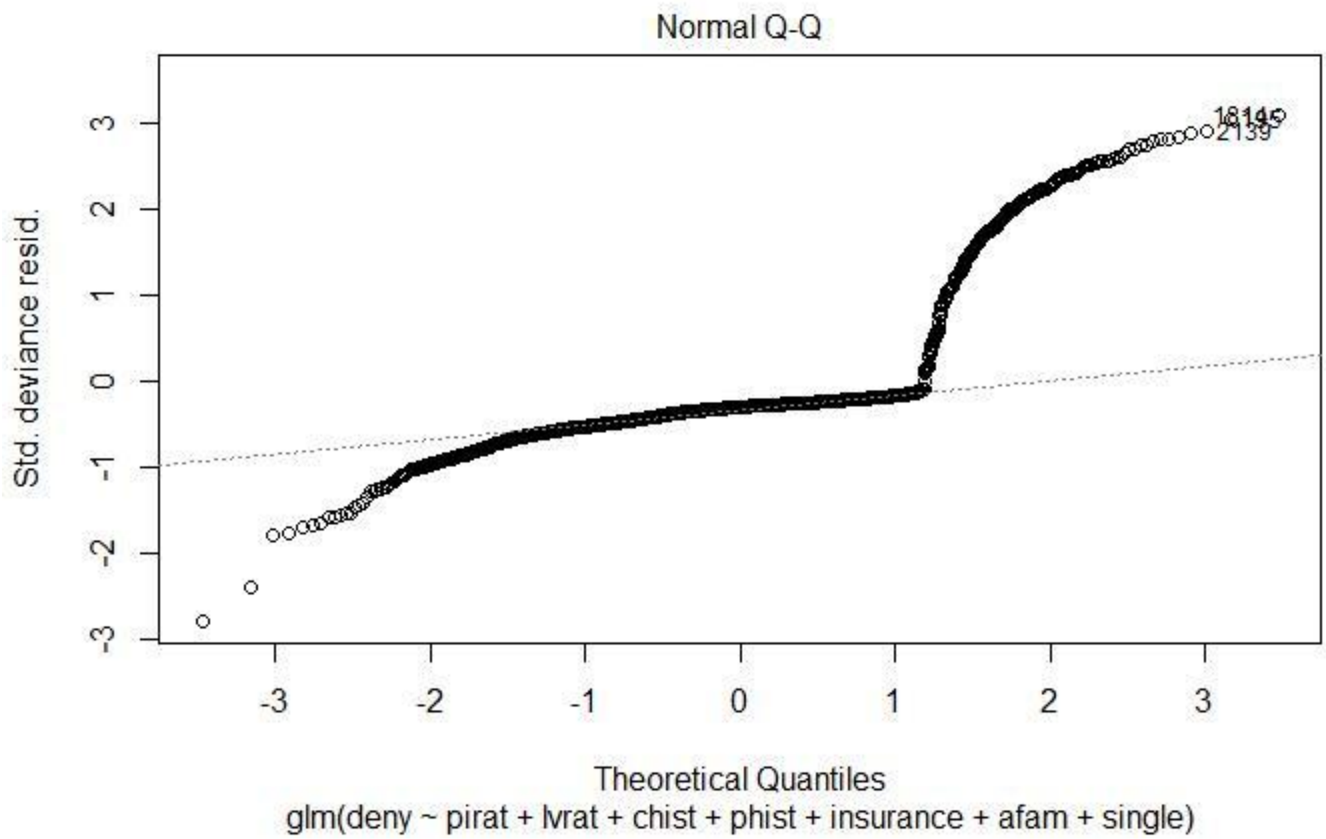
```
fm3 <- lm(I(as.numeric(deny) - 1) ~., data = dataset)
summary(fm3)
```

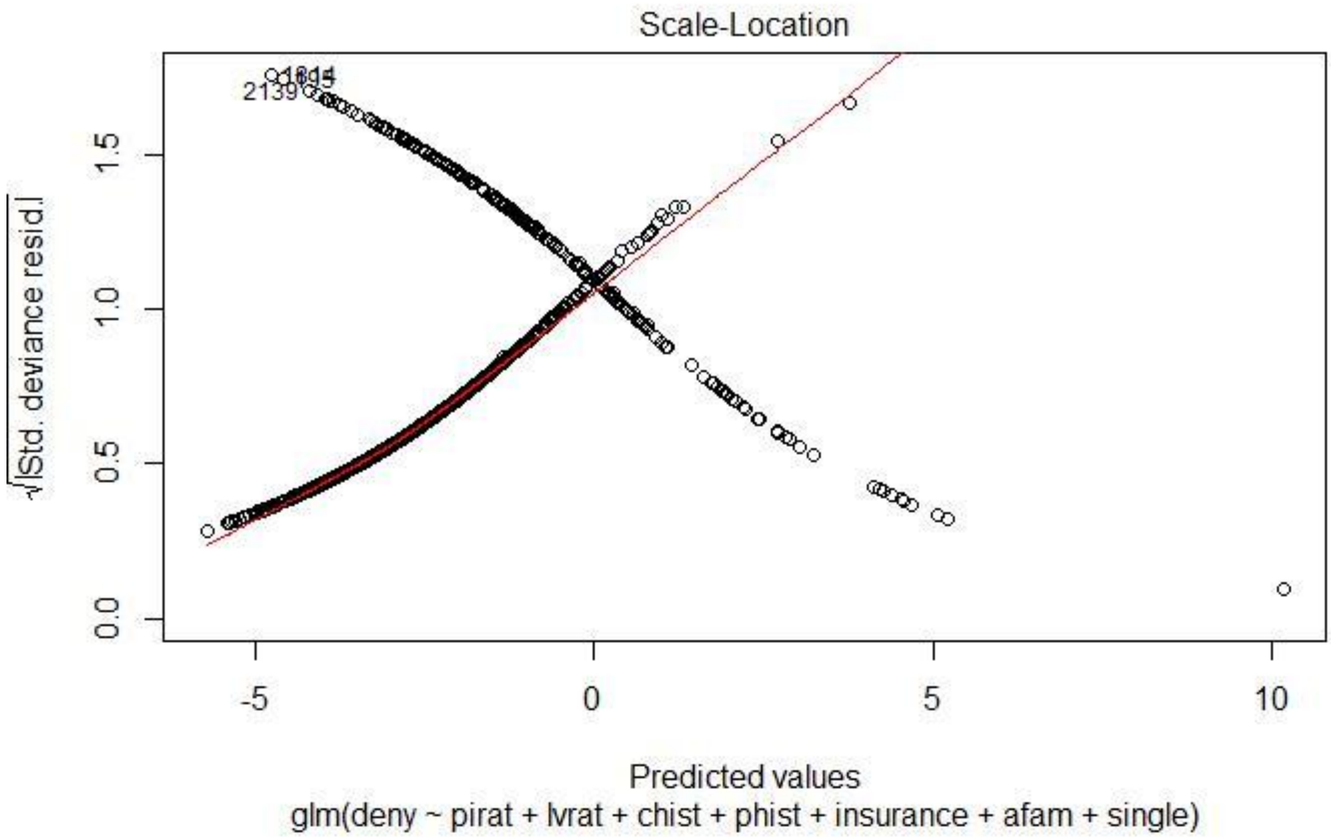
```
Call:
lm(formula = I(as.numeric(deny) - 1) ~ ., data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.92559 -0.12214 -0.05246 -0.00258 1.07741 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.095597   0.057354  -1.667  0.09569 .  
pirat         0.476759   0.088653   5.378 8.28e-08 ***
hirat        -0.099505   0.097994  -1.015  0.31001    
lvrat         0.094726   0.033942   2.791  0.00530 ** 
chist2        0.038143   0.015483   2.464  0.01383 *  
chist3        0.049182   0.026661   1.845  0.06520 .  
chist4        0.135810   0.033372   4.070 4.86e-05 ***
chist5        0.106922   0.022466   4.759 2.06e-06 ***
chist6        0.160628   0.022773   7.054 2.28e-12 ***
mhist2        0.018115   0.013432   1.349  0.17758    
mhist3        0.034372   0.045616   0.754  0.45122    
mhist4        0.027747   0.062691   0.443  0.65810    
phistyes      0.204367   0.023531   8.685  < 2e-16 ***
unemp         0.004688   0.002903   1.615  0.10650    
selfempyes    0.056660   0.018305   3.095  0.00199 ** 
insuranceyes  0.713226   0.041360  17.244  < 2e-16 ***
condominyes  -0.005913   0.013654  -0.433  0.66501    
afamyes       0.084846   0.017537   4.838 1.40e-06 ***
singleyes     0.035722   0.012599   2.835  0.00462 ** 
hschoolyes   -0.115475   0.045984  -2.511  0.01210 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2796 on 2360 degrees of freedom
Multiple R-squared:  0.2643,	Adjusted R-squared:  0.2584 
F-statistic: 44.63 on 19 and 2360 DF,  p-value: < 2.2e-16
```

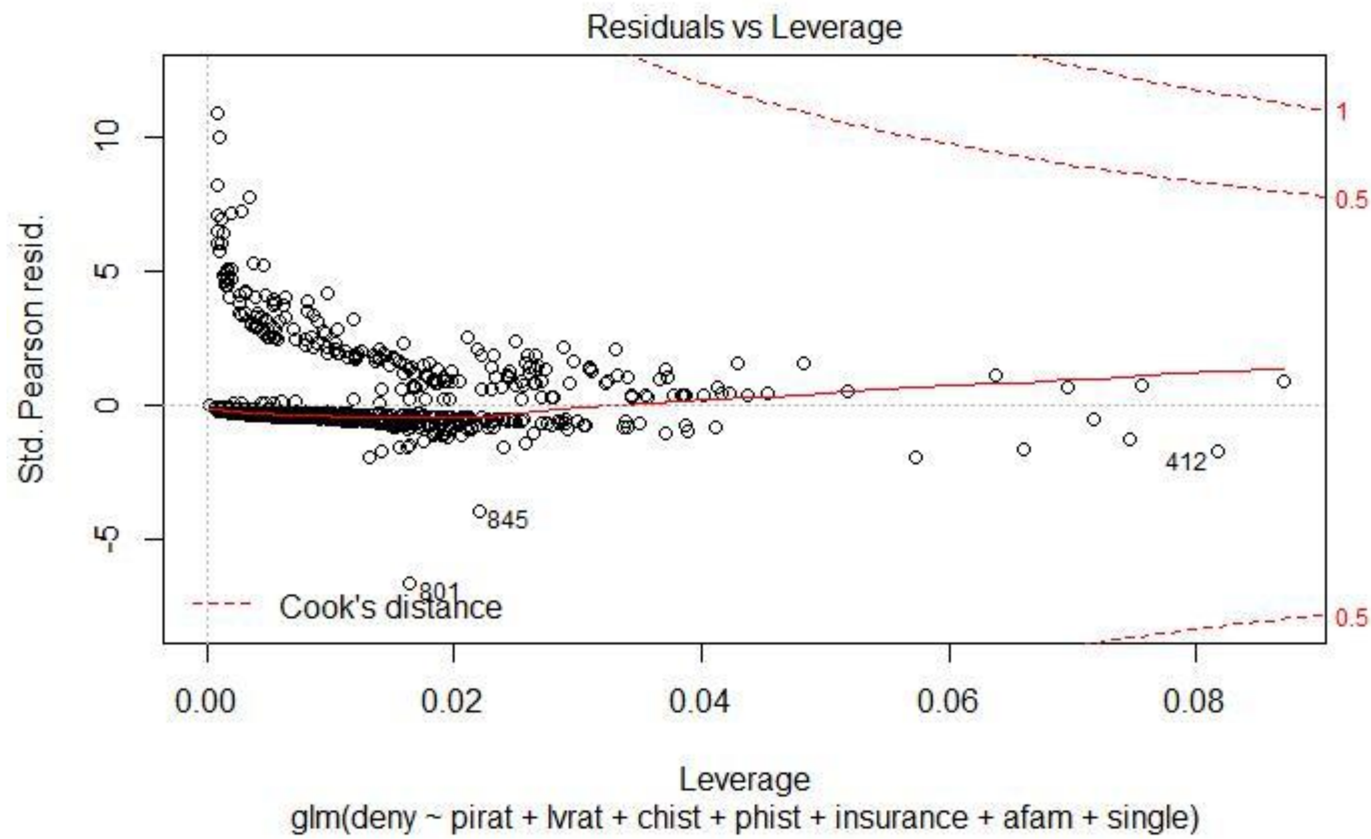
Hide

```
plot(model3)
```


Residuals vs Fitted
Residuals
1814
2139
-5
0
5
10
-3
-2
-1
0
1
2
3
Predicted values
glm(deny ~ pirat + lvrat + chist + phist + insurance + afam + single)


Normal Q-Q
Std. deviance resid.
1814
2139
-3
-2
-1
0
1
2
3
Theoretical Quantiles
glm(deny ~ pirat + lvrat + chist + phist + insurance + afam + single)





Partial F-Test

Hide

```
anova(full_model,model3)
```

```
Analysis of Deviance Table

Model 1: deny ~ pirat + hirat + lvrat + chist + mhist + phist + unemp
    + selfemp + insurance + condomin + afam + single + hschool +
    model3
Model 2: deny ~ pirat + lvrat + chist + phist + insurance + afam +
  single Resid. Df Resid. Dev Df Deviance
1      1883     993.08
2      18921015.18 -9  -22.107
```

Assessing the predictive ability of the model

Hide

```
train$model3 <- predict(model2, train, type="response")
head(train)
```

	deny <fctr>	pirat <dbl>	hirat <dbl>	lvrat <dbl>	chist <fctr>	mhist <fctr>	phist <fctr>	unemp <dbl>	selfemp <fctr>
1	no	0.221	0.221	0.8000000	5	2	no	3.9	no
2	no	0.265	0.265	0.9218750	2	2	no	3.2	no
3	no	0.372	0.248	0.9203980	1	2	no	3.2	no
4	no	0.320	0.250	0.8604651	1	2	no	4.3	no
5	no	0.360	0.350	0.6000000	1	1	no	3.2	no
6	no	0.240	0.170	0.5105263	1	1	no	3.9	no

6 rows | 1-10 of 15 columns

Hide

```
tail(train)
```

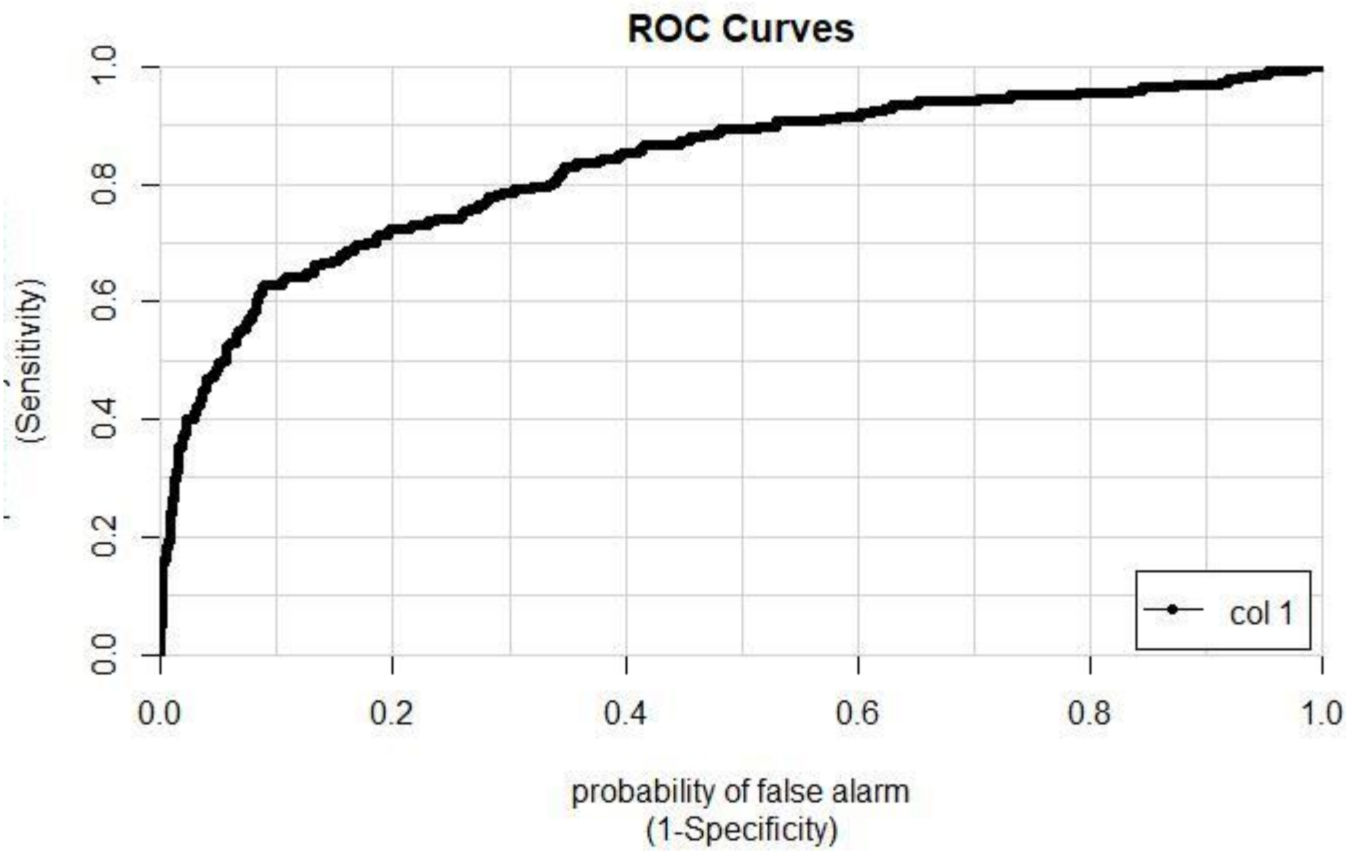
	deny <fctr>	pirat <dbl>	hirat <dbl>	lvrat <dbl>	chist <fctr>	mhist <fctr>	phist <fctr>	unemp <dbl>	selfemp <fctr>
2374	no	0.35	0.22	0.8939394	3	2	no	3.9	no
2376	no	0.31	0.25	0.8000000	1	1	no	3.2	yes
2377	no	0.30	0.30	0.7770492	1	2	no	3.2	no
2378	no	0.26	0.20	0.5267606	2	1	no	3.1	no
2379	yes	0.32	0.26	0.7538462	6	1	yes	3.1	no
2380	yes	0.35	0.26	0.8135593	2	2	no	4.3	no

6 rows | 1-10 of 15 columns

Hide

```
library(gmodels)
library(ggplot2)
library (Hmisc)
library (caTools)
library (ROCR)
colAUC(train$model3,train$deny, plotROC=TRUE)
```

```
                [,1]
no vs. yes 0.8316539
```



Hide

```
predict1 <- ifelse(train$model3>0.84, 1, 0)
predict1
```

```
   [1] 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
  [31] 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  [61] 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  [91] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [121] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [151] 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [181] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [211] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [241] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [271] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [301] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [331] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [361] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [391] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
 [421] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
 [451] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [481] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [511] 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
 [541] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [571] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [601] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 1 1 1
 [631] 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
 [661] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [691] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [721] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [751] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [781] 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [811] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [841] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [871] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [901] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [931] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [961] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 [991] 0 0 0 0 0 0 0 0 0 0
 [ reached getOption("max.print") -- omitted 904 entries ]
```

Hide

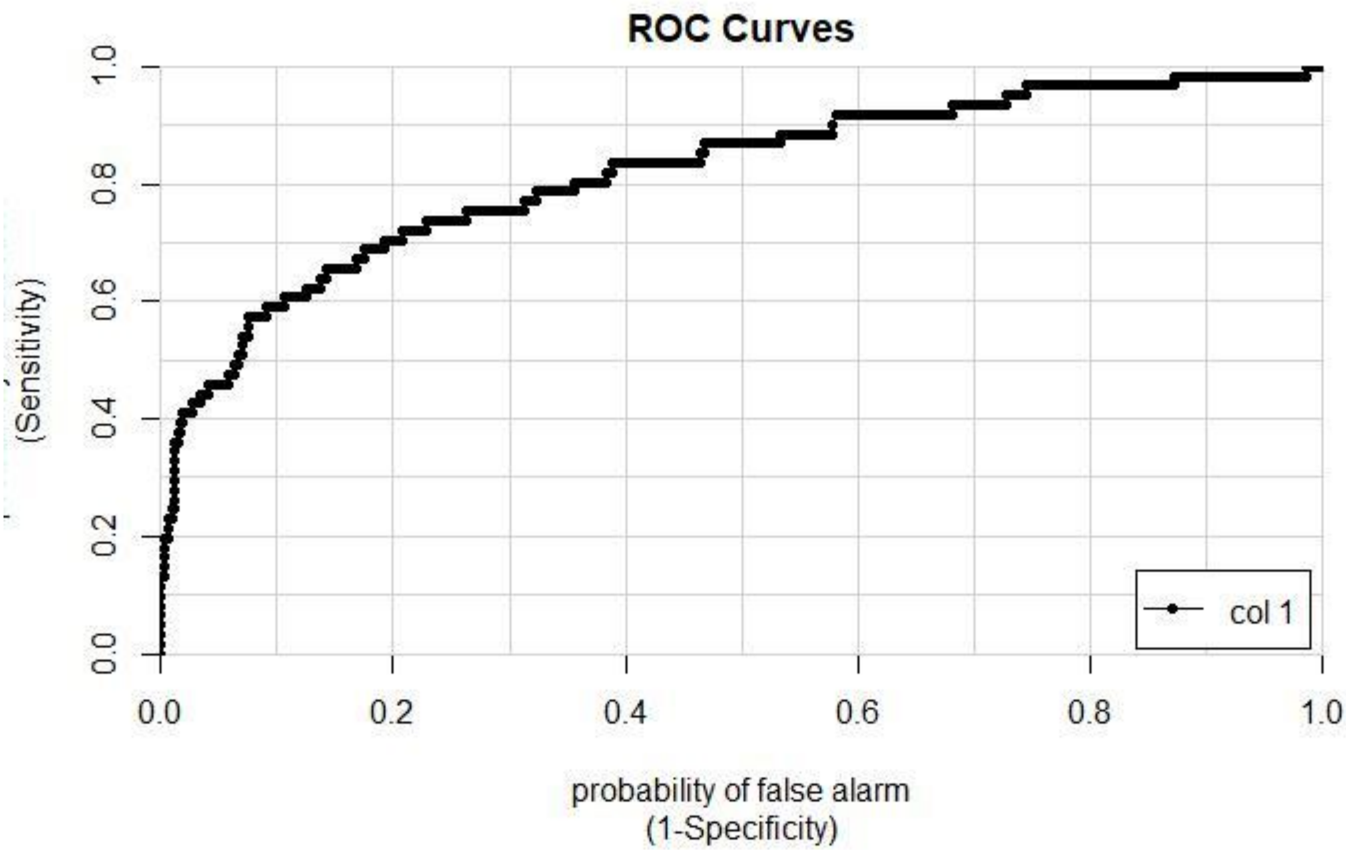
```
tab1 <- table(predicted = predict1, actual = train$deny)
tab1
```

```
         actual
predicted   no  yes
0 1678                                        192
        1    2   32
```

Hide

```
test$model3 <- predict(model3, test, type='response')
colAUC(test$model3,test$deny, plotROC=TRUE)
```

```
                [,1]
no vs. yes 0.8199289
```



Hide

```
predict2 <- ifelse(test$model3>0.78, 1,0)
tab2 <- table(predicted = predict2, actual = test$deny)
tab2
```

```
         actual
predicted  no yes
        0 413  49
        1   2  12
```

CONCLUSION: - From the above analysis we can conclude that below are the most significant predictors on which Home Mortgage acceptance or denied depends.

- Payment to Income Ratio
- Loan to value ratio
- Credit history: consumer payments
- Public bad credit record
- Insurance
- Ethnicity
- Marital status