Home Mortgage Disclosure Act Data

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)))
```

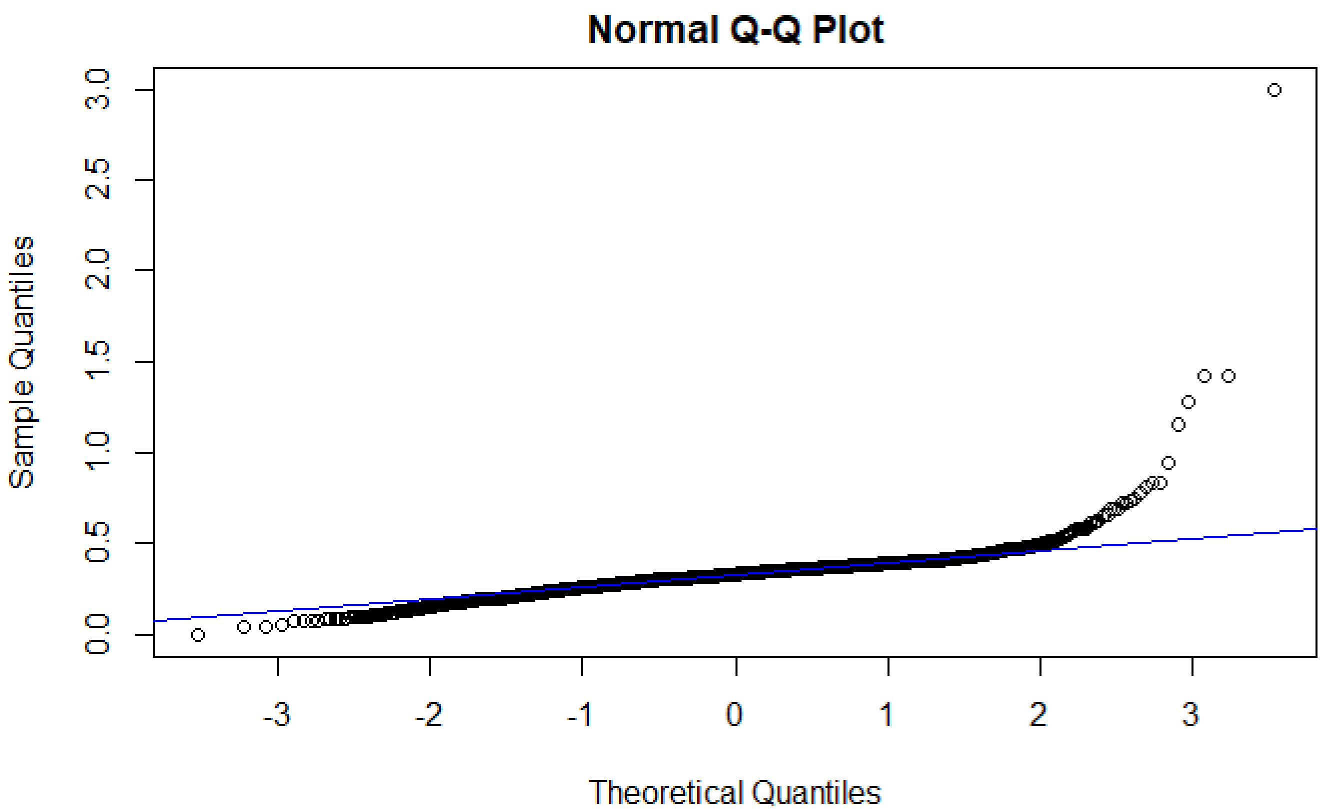
```
     deny     pirat     hirat     lvrat     chist     mhist 
        0         0         0         0         0         0 
    phist     unemp   selfemp insurance  condomin      afam 
        0         0         0         0         0         0 
   single   hschool 
        0         0 
```

As we can see data is 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

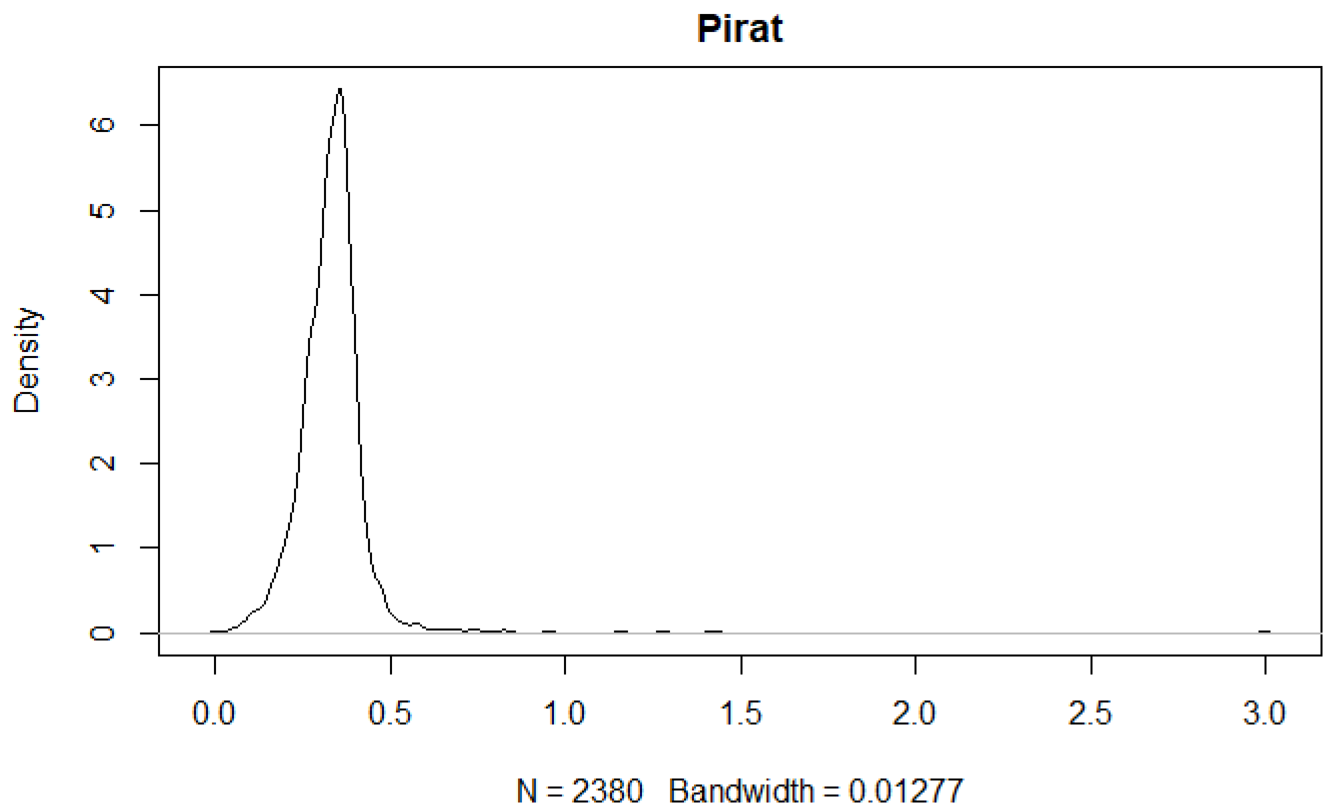
Hide

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



Hide

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

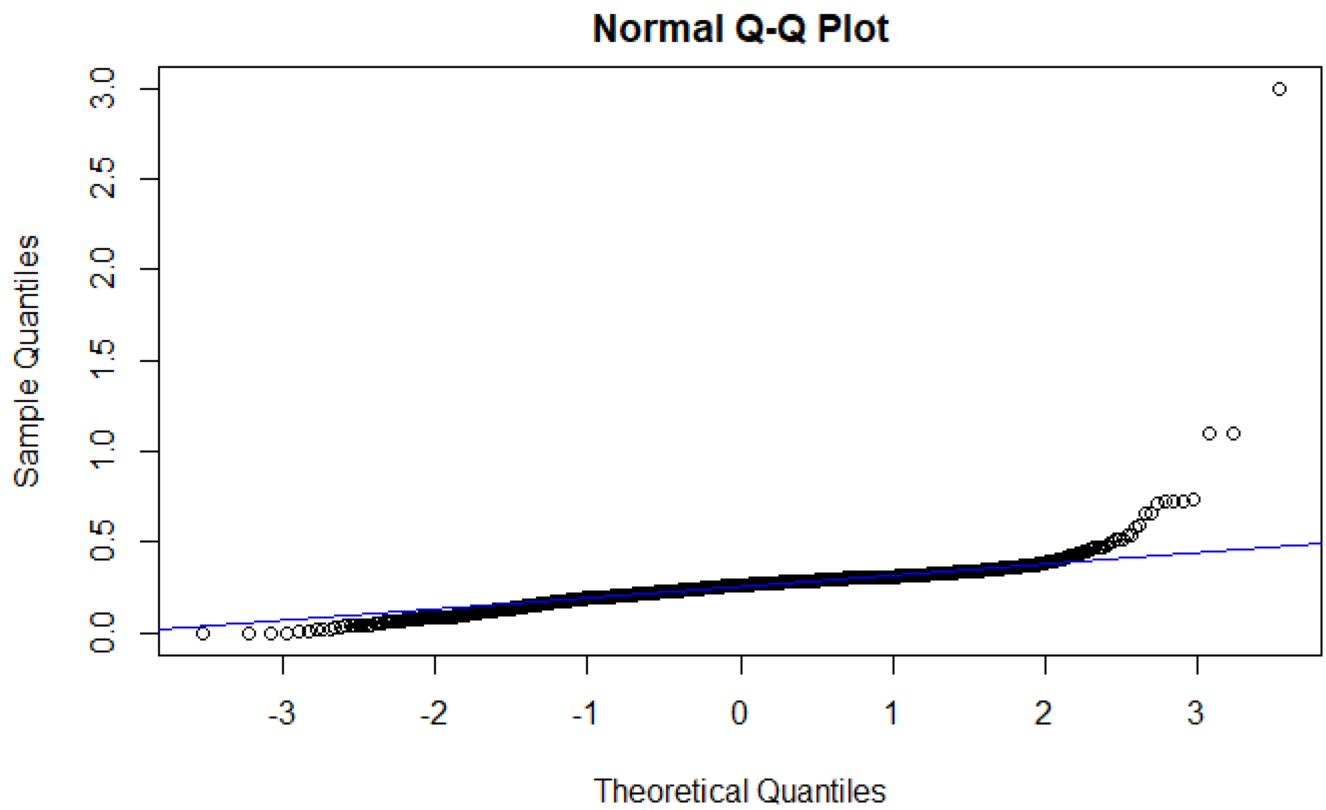


The plot looks right skewed with few outliers.

QQ-plot and density plots housing expense to income ratio

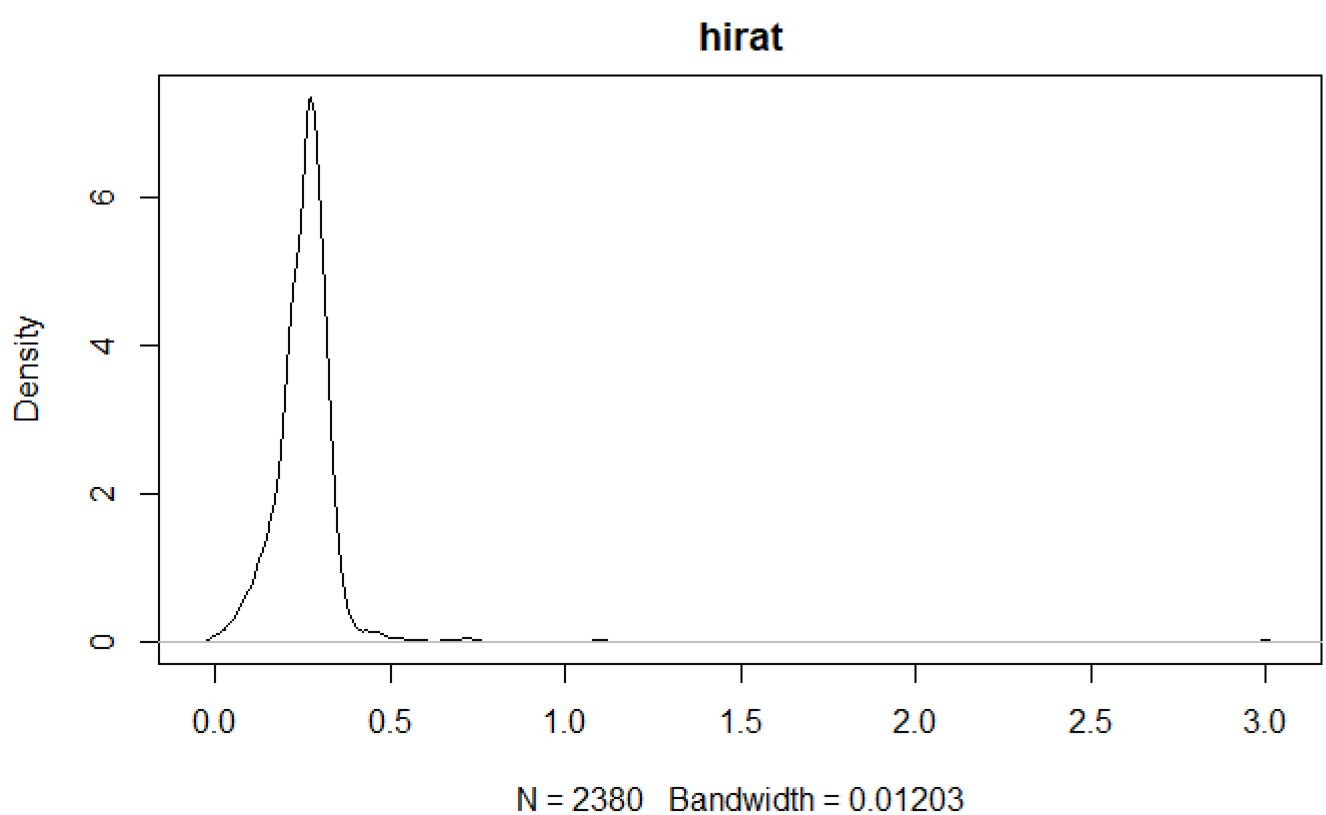
Hide

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



Hide

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

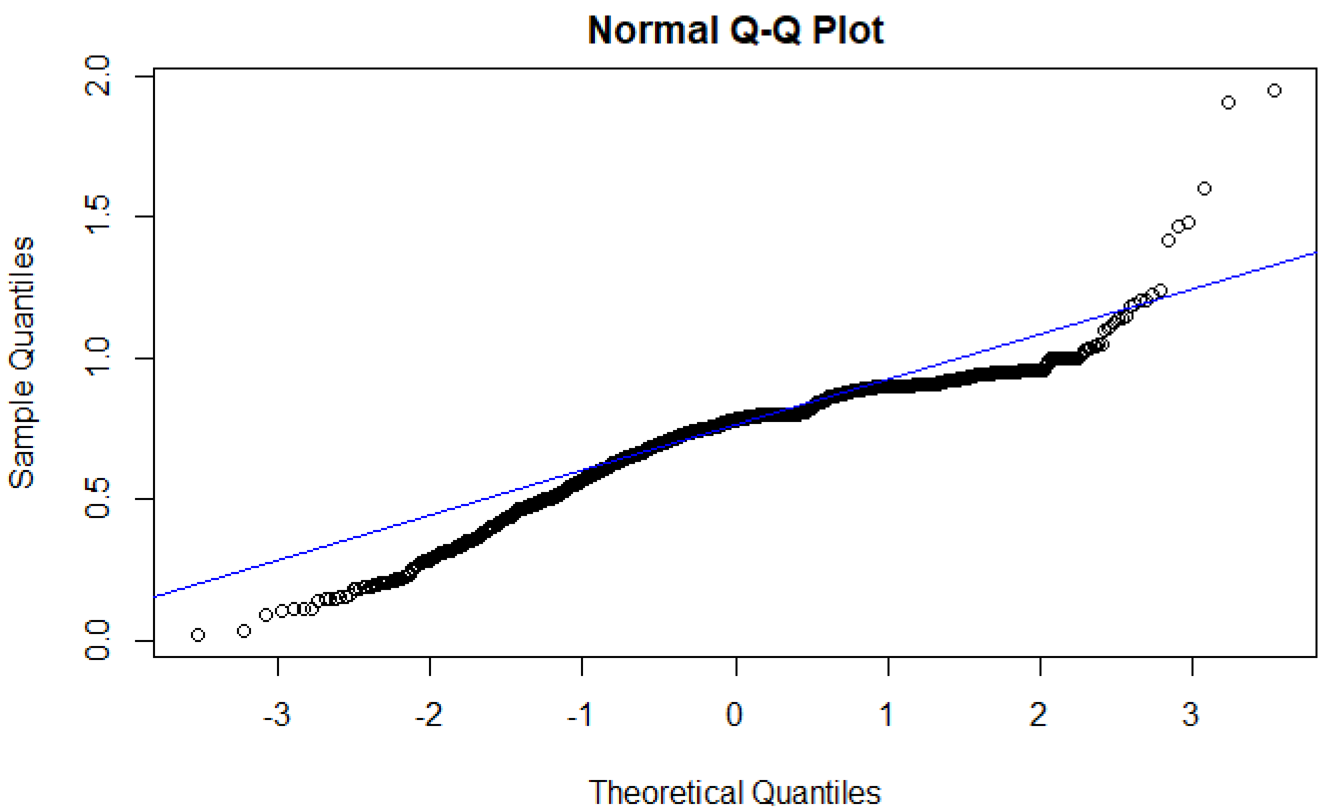


The plot looks right skewed with few outliers.

QQ-plot and dnsity plots Loan to value ratio

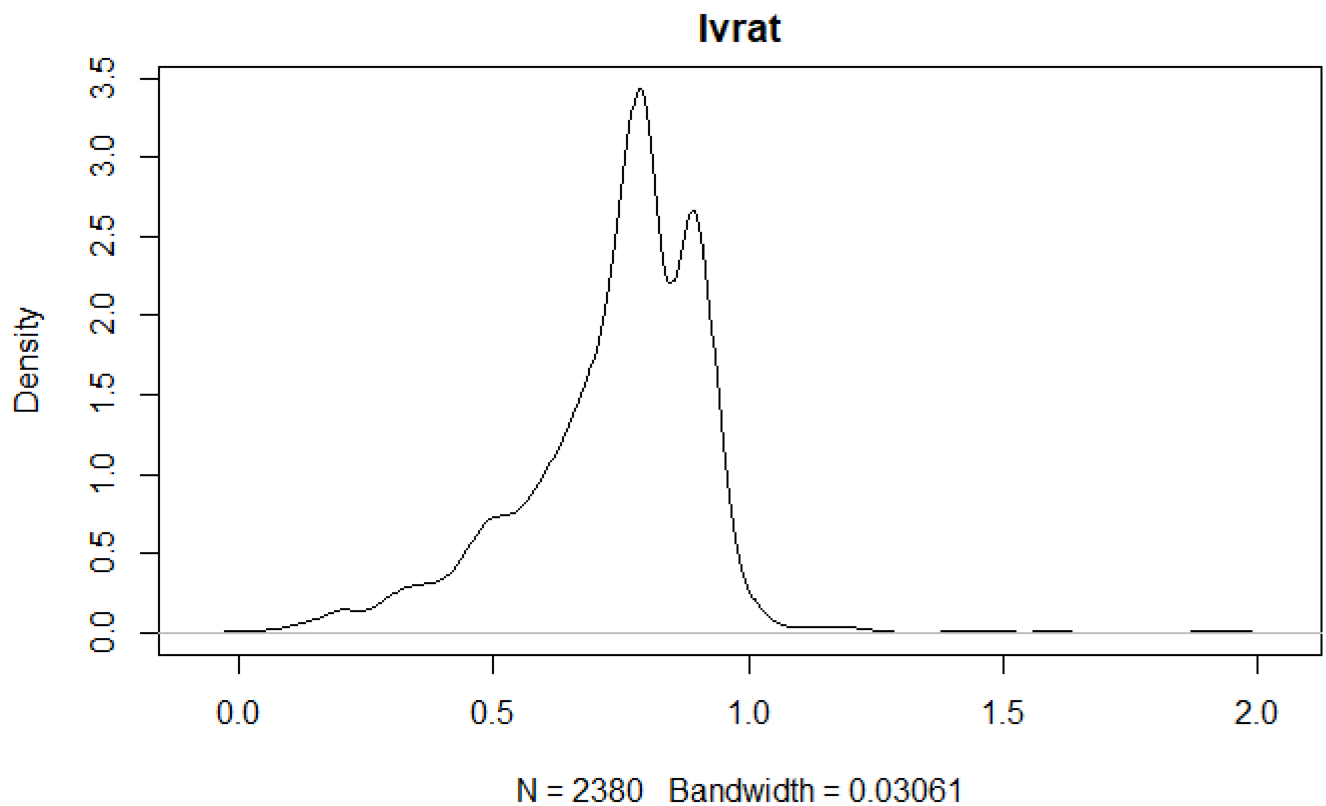
Hide

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



Hide

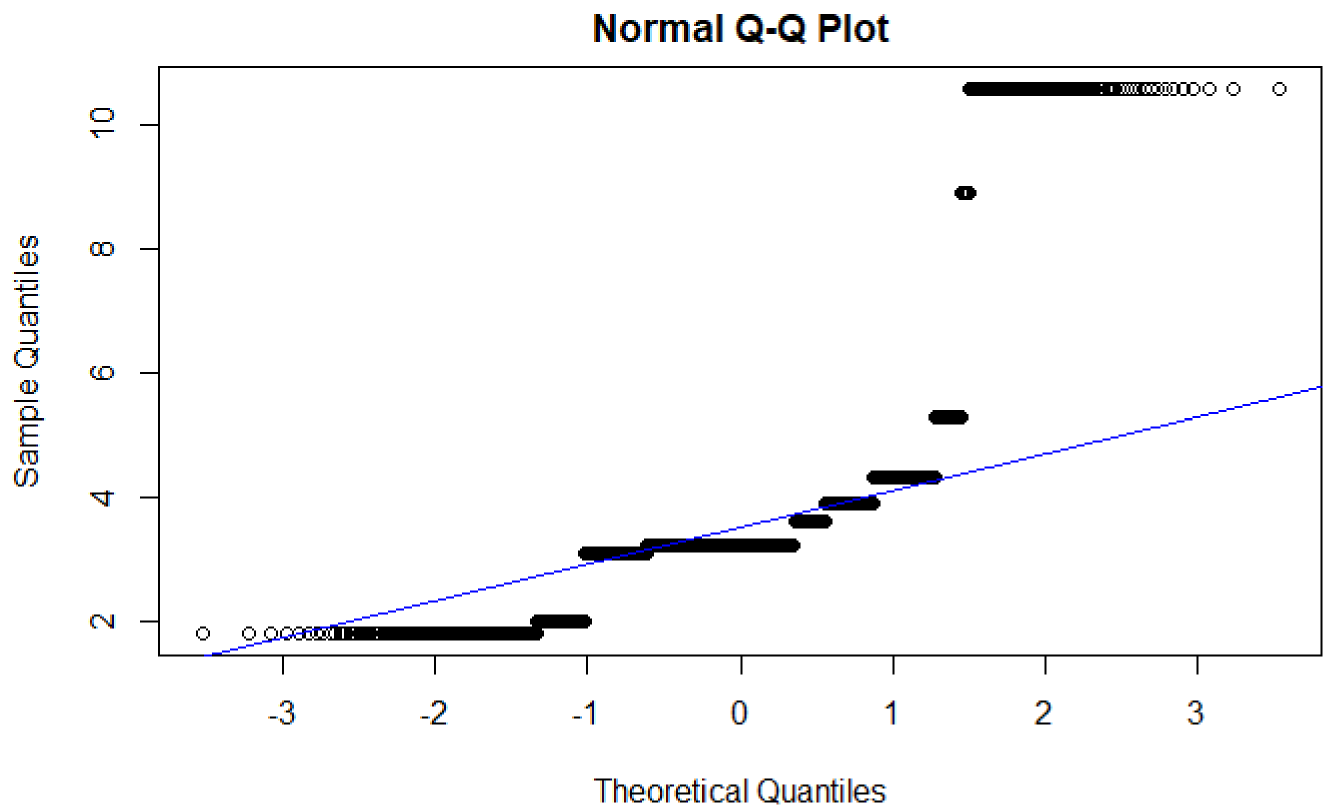
```
plot(density(dataset$lvrat),main='lvrat')
```



QQ-plot and density plots unemployment

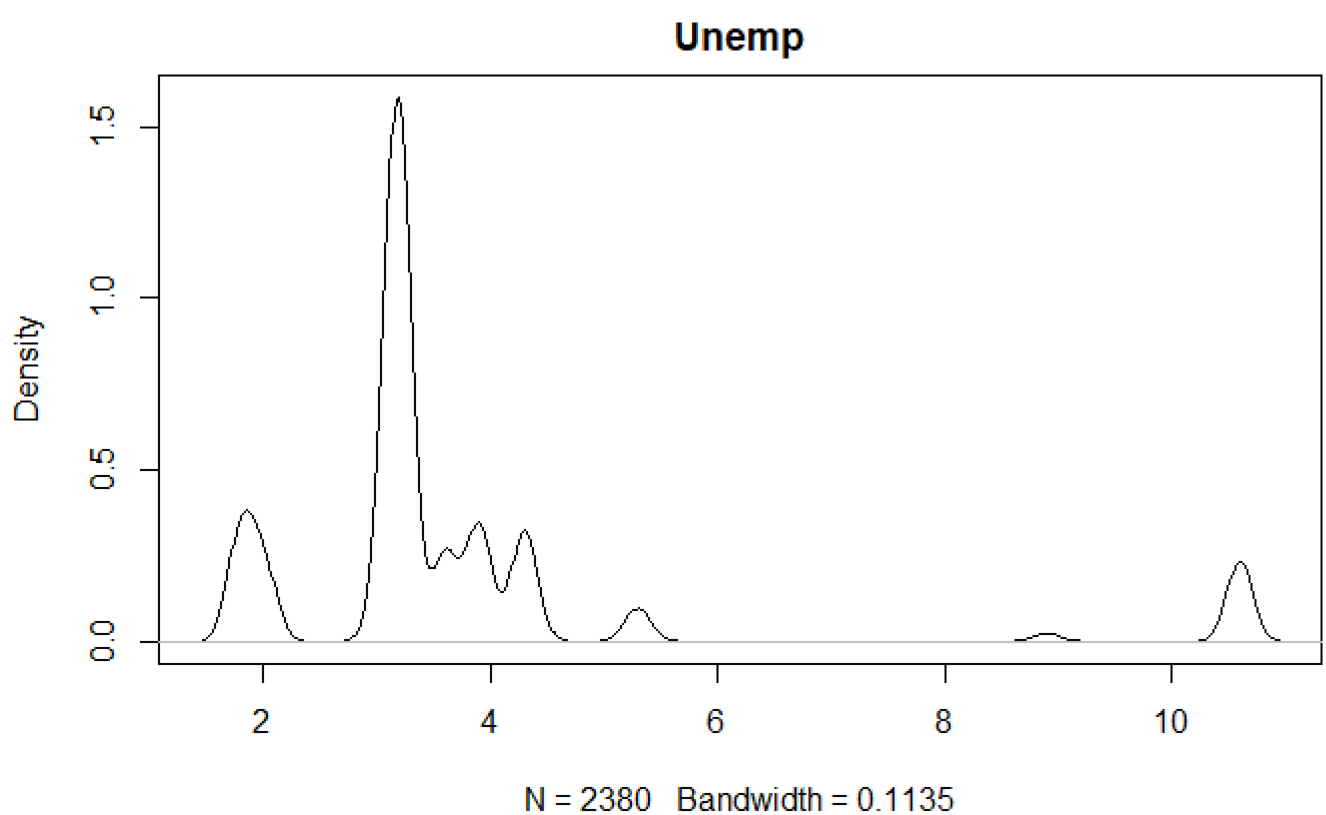
Hide

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



Hide

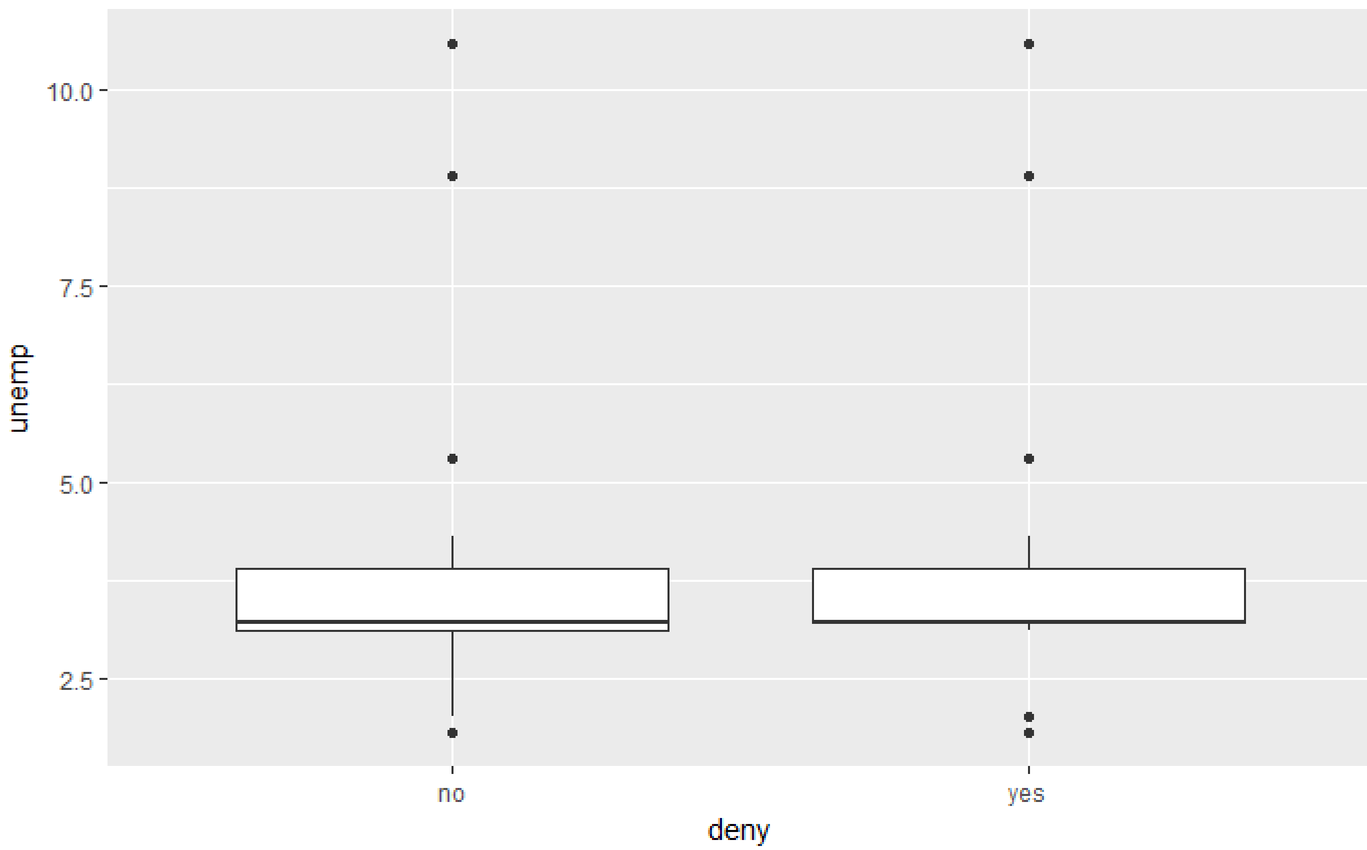
```
plot(density(dataset$unemp),main='Unemp')
```



Box-Plots

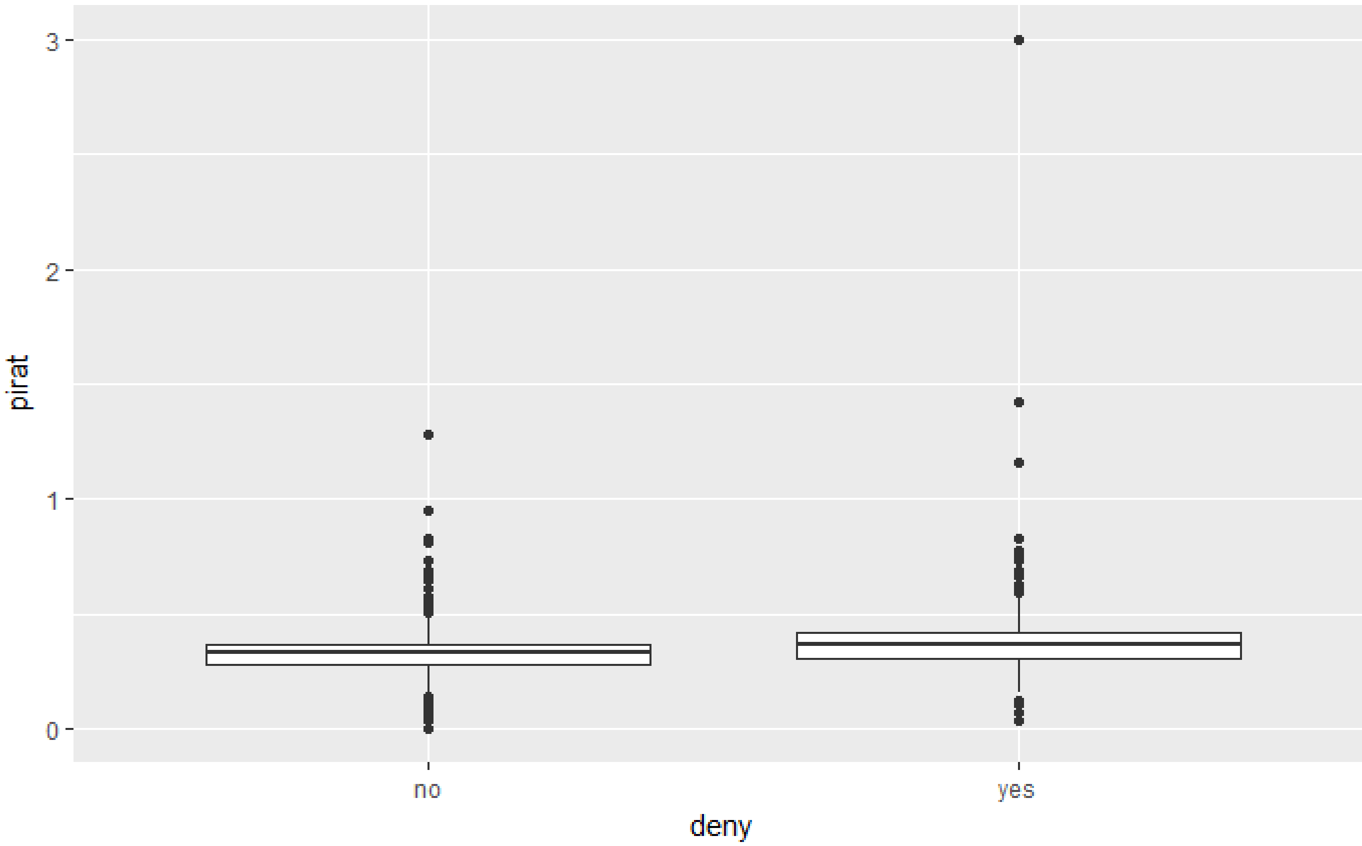
Hide

```
library(ggplot2)
p10 <- ggplot(dataset, aes(x = deny, y = unemp)) +
       geom_boxplot()
p10
```



Hide

```
p11 <- ggplot(dataset, aes(x = deny, y = pirat)) +
       geom_boxplot()
p11
```



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

Model fitting:

We split the data into two chunks: 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  
-2.7851  -0.4209  -0.2986  -0.2052   3.0703  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -6.06081    0.80635  -7.516 5.63e-14 ***
pirat         5.68727    1.26257   4.505 6.65e-06 ***
hirat        -0.98954    1.43543  -0.689 0.490590    
lvrat         2.16966    0.60116   3.609 0.000307 ***
chist2        0.58326    0.23845   2.446 0.014445 *  
chist3        0.65820    0.36733   1.792 0.073154 .  
chist4        1.41414    0.39140   3.613 0.000303 ***
chist5        1.14123    0.27684   4.122 3.75e-05 ***
chist6        1.51969    0.25950   5.856 4.74e-09 ***
mhist2        0.25891    0.22035   1.175 0.239992    
mhist3        0.03562    0.54847   0.065 0.948225    
mhist4        0.10461    0.85091   0.123 0.902152    
phistyes      1.42219    0.23797   5.976 2.28e-09 ***
unemp         0.06084    0.03990   1.525 0.127331    
selfempyes    0.80705    0.24182   3.337 0.000846 ***
insuranceyes  4.43250    0.56881   7.793 6.57e-15 ***
condominyes  -0.07214    0.19511  -0.370 0.711564    
afamyes       0.52060    0.20872   2.494 0.012621 *  
singleyes     0.33171    0.17982   1.845 0.065089 .  
hschoolyes   -1.03454    0.48473  -2.134 0.032820 *  
---
Signif. codes:  0 ***  0.001 **  0.01 *  0.05 .  0.1    1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.10  on 1903  degrees of freedom
Residual deviance:  993.29  on 1884  degrees of freedom
AIC: 1033.3

Number of Fisher Scoring iterations: 6
```

By using function summary() we obtain the results of our model. Interpreting the results of our logistic regression model: Now we can analyze the fitting and interpret what the model is telling us. First of all, 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    1363.10              
pirat      1   62.871      1902    1300.23 2.207e-15 ***
hirat      1    1.248      1901    1298.98  0.263860    
lvrat      1   49.191      1900    1249.79 2.322e-12 ***
chist      5   84.696      1895    1165.09 < 2.2e-16 ***
mhist      3    4.496      1892    1160.59  0.212618    
phist      1   34.695      1891    1125.90 3.856e-09 ***
unemp      1    4.365      1890    1121.53  0.036691 *  
selfemp    1    9.271      1889    1112.26  0.002329 ** 
insurance  1  105.444      1888    1006.82 < 2.2e-16 ***
condomin   1    0.378      1887    1006.44  0.538540    
afam       1    6.488      1886     999.95  0.010859 *  
single     1    2.595      1885     997.36  0.107196    
hschool    1    4.070      1884     993.29  0.043640 *  
---
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  
-2.9276  -0.4199  -0.3009  -0.2154   3.1311  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -5.6749     0.7472  -7.595 3.07e-14 ***
pirat          5.0779     0.9352   5.430 5.65e-08 ***
lvrat          2.2599     0.5949   3.799 0.000146 ***
chist2         0.5991     0.2366   2.532 0.011349 *  
chist3         0.6451     0.3628   1.778 0.075351 .  
chist4         1.3837     0.3882   3.564 0.000365 ***
chist5         1.1493     0.2756   4.170 3.05e-05 ***
chist6         1.5505     0.2544   6.096 1.09e-09 ***
phistyes       1.4151     0.2373   5.963 2.47e-09 ***
selfempyes     0.8260     0.2370   3.485 0.000492 ***
insuranceyes   4.4635     0.5661   7.884 3.16e-15 ***
afamyes        0.5115     0.2048   2.498 0.012494 *  
singleyes      0.3339     0.1721   1.940 0.052357 .  
hschoolyes    -1.1515     0.4782  -2.408 0.016028 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.10  on 1903  degrees of freedom
Residual deviance:  997.53  on 1890  degrees of freedom
AIC: 1025.5

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    1363.10              
pirat      1   62.871      1902    1300.23 2.207e-15 ***
lvrat      1   48.819      1901    1251.41 2.807e-12 ***
chist      5   86.015      1896    1165.39 < 2.2e-16 ***
phist      1   34.562      1895    1130.83 4.128e-09 ***
selfemp    1    9.172      1894    1121.66  0.002458 ** 
insurance  1  108.671      1893    1012.99 < 2.2e-16 ***
afam       1    7.347      1892    1005.64  0.006716 ** 
single     1    3.051      1891    1002.59  0.080708 .  
hschool    1    5.057      1890     997.53  0.024533 *  
---
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  
-2.9383  -0.4263  -0.3119  -0.2219   3.0687  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.5964     0.5863 -11.252  < 2e-16 ***
pirat          5.0623     0.9326   5.428 5.70e-08 ***
lvrat          2.1663     0.5957   3.636 0.000277 ***
chist2         0.6378     0.2349   2.715 0.006637 ** 
chist3         0.6854     0.3578   1.916 0.055385 .  
chist4         1.4107     0.3803   3.710 0.000207 ***
chist5         1.2025     0.2736   4.395 1.11e-05 ***
chist6         1.5011     0.2524   5.947 2.73e-09 ***
phistyes       1.4213     0.2358   6.027 1.67e-09 ***
insuranceyes   4.4289     0.5666   7.816 5.45e-15 ***
afamyes        0.4951     0.2030   2.439 0.014717 *  
singleyes      0.2901     0.1698   1.708 0.087626 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1013.1  on 1892  degrees of freedom
AIC: 1037.1

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     1363.1
pirat      1   62.871      1902     1300.2 2.207e-15 ***
lvrat      1   48.819      1901     1251.4 2.807e-12 ***
chist      5   86.015      1896     1165.4 < 2.2e-16 ***
phist      1   34.562      1895     1130.8 4.128e-09 ***
insurance  1  108.191      1894     1022.6 < 2.2e-16 ***
afam       1    6.590      1893     1016.0   0.01026 *
single     1    2.902      1892     1013.1   0.08846 .
---
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  
-1.9669  -0.5198  -0.4617  -0.3756   2.8163  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -4.1984     0.3156 -13.304  < 2e-16 ***
pirat         6.2911     0.8644   7.278 3.39e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1300.2  on 1902  degrees of freedom
AIC: 1304.2

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  
-2.0510  -0.5246  -0.4557  -0.3771   2.7859  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -4.1006     0.3254 -12.603  < 2e-16 ***
pirat         6.9987     1.0654   6.569 5.06e-11 ***
hirat        -1.3066     1.1607  -1.126     0.26    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1299.0  on 1901  degrees of freedom
AIC: 1305

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
-1.7020  -0.5453  -0.4289  -0.2914   3.1713

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.7419     0.5336 -12.635  < 2e-16 ***
pirat         6.0314     0.8813   6.844 7.70e-12 ***
lvrat         3.3823     0.5204   6.500 8.04e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1251.4  on 1901  degrees of freedom
AIC: 1257.4

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
-1.9901  -0.4863  -0.3655  -0.2411   3.3151

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.0869     0.5529 -12.817  < 2e-16 ***
pirat         5.5324     0.9030   6.127 8.96e-10 ***
lvrat         3.2008     0.5350   5.983 2.19e-09 ***
chist2        0.6752     0.2134   3.164 0.001555 **
chist3        0.9876     0.3140   3.145 0.001660 **
chist4        1.2930     0.3697   3.498 0.000469 ***
chist5        1.2534     0.2489   5.035 4.77e-07 ***
chist6        1.9118     0.2175   8.788  < 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: 1363.1  on 1903  degrees of freedom
Residual deviance: 1165.4  on 1896  degrees of freedom
AIC: 1181.4

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
-1.9658  -0.4863  -0.3642  -0.2350   3.2529

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.2324     0.5614 -12.882  < 2e-16 ***
pirat         5.5719     0.9136   6.099 1.07e-09 ***
lvrat         3.0173     0.5398   5.590 2.27e-08 ***
chist2        0.6913     0.2137   3.234 0.001219 **
chist3        1.0103     0.3180   3.177 0.001488 **
chist4        1.2378     0.3710   3.337 0.000848 ***
chist5        1.2469     0.2491   5.006 5.55e-07 ***
chist6        1.8603     0.2206   8.432  < 2e-16 ***
mhist2        0.3782     0.1949   1.941 0.052266 .
mhist3        0.2038     0.5271   0.387 0.698992
mhist4        0.4431     0.7276   0.609 0.542584
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1161.4  on 1893  degrees of freedom
AIC: 1183.4

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
-1.7269  -0.4704  -0.3540  -0.2379   3.3042

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.9880     0.5574 -12.538  < 2e-16 ***
pirat         5.2821     0.9100   5.804 6.46e-09 ***
lvrat         3.1080     0.5388   5.768 8.01e-09 ***
chist2        0.6188     0.2157   2.868  0.00413 **
chist3        0.7358     0.3254   2.261  0.02377 *
chist4        1.3161     0.3711   3.547  0.00039 ***
chist5        1.1144     0.2554   4.363 1.29e-05 ***
chist6        1.5057     0.2345   6.420 1.36e-10 ***
phistyes      1.3638     0.2248   6.066 1.31e-09 ***
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1130.8  on 1895  degrees of freedom
AIC: 1148.8

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
-1.9589  -0.4668  -0.3497  -0.2328   3.3526

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.30941    0.58100 -12.581  < 2e-16 ***
pirat        5.26079    0.91344   5.759 8.44e-09 ***
lvrat        3.14372    0.53943   5.828 5.61e-09 ***
chist2       0.59518    0.21641   2.750 0.005954 **
chist3       0.72200    0.32565   2.217 0.026618 *
chist4       1.34082    0.37164   3.608 0.000309 ***
chist5       1.11903    0.25611   4.369 1.25e-05 ***
chist6       1.52180    0.23495   6.477 9.35e-11 ***
phistyes     1.36543    0.22451   6.082 1.19e-09 ***
unemp        0.07829    0.03580   2.187 0.028768 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1126.4  on 1894  degrees of freedom
AIC: 1146.4

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  
-2.0211  -0.4652  -0.3503  -0.2340   3.3500  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -7.1662     0.5619 -12.752  < 2e-16 ***
pirat         5.3239     0.9122   5.836 5.35e-09 ***
lvrat         3.2009     0.5427   5.899 3.66e-09 ***
chist2        0.6025     0.2166   2.782 0.005405 ** 
chist3        0.7264     0.3261   2.227 0.025921 *  
chist4        1.3506     0.3733   3.618 0.000297 ***
chist5        1.0844     0.2576   4.210 2.56e-05 ***
chist6        1.5488     0.2357   6.572 4.96e-11 ***
phistyes      1.3602     0.2248   6.051 1.44e-09 ***
selfempyes    0.7121     0.2256   3.156 0.001597 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1121.7  on 1894  degrees of freedom
AIC: 1141.7

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
-2.9240  -0.4257  -0.3192  -0.2250   3.2473

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.6588     0.5818 -11.445  < 2e-16 ***
pirat          5.2216     0.9366   5.575 2.47e-08 ***
lvrat          2.4085     0.5871   4.103 4.09e-05 ***
chist2         0.6490     0.2342   2.771  0.00559 **
chist3         0.7494     0.3534   2.121  0.03395 *
chist4         1.5678     0.3752   4.179 2.93e-05 ***
chist5         1.2566     0.2693   4.666 3.08e-06 ***
chist6         1.6175     0.2476   6.533 6.44e-11 ***
phistyes       1.4524     0.2323   6.252 4.05e-10 ***
insuranceyes   4.4254     0.5653   7.828 4.95e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1022.6  on 1894  degrees of freedom
AIC: 1042.6

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     1363.1
pirat      1   62.871      1902     1300.2 2.207e-15 ***
lvrat      1   48.819      1901     1251.4 2.807e-12 ***
chist      5   86.015      1896     1165.4 < 2.2e-16 ***
phist      1   34.562      1895     1130.8 4.128e-09 ***
insurance  1  108.191      1894     1022.6 < 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  
-2.9458  -0.4241  -0.3190  -0.2261   3.2216  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.6743     0.5819 -11.471  < 2e-16 ***
pirat          5.2246     0.9353   5.586 2.32e-08 ***
lvrat          2.3880     0.5877   4.063 4.85e-05 ***
chist2         0.6573     0.2348   2.799  0.00512 ** 
chist3         0.7495     0.3532   2.122  0.03386 *  
chist4         1.5438     0.3774   4.090 4.30e-05 ***
chist5         1.2546     0.2694   4.656 3.22e-06 ***
chist6         1.6093     0.2481   6.488 8.72e-11 ***
phistyes       1.4519     0.2324   6.247 4.17e-10 ***
insuranceyes   4.4344     0.5658   7.837 4.60e-15 ***
condominyes    0.1023     0.1813   0.564  0.57250    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1022.3  on 1893  degrees of freedom
AIC: 1044.3

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  
-2.9968  -0.4263  -0.3163  -0.2259   3.1132  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.4904     0.5813 -11.165  < 2e-16 ***
pirat          5.0777     0.9348   5.432 5.57e-08 ***
lvrat          2.1739     0.5937   3.662 0.000251 ***
chist2         0.6372     0.2345   2.717 0.006593 ** 
chist3         0.7061     0.3550   1.989 0.046708 *  
chist4         1.4598     0.3790   3.852 0.000117 ***
chist5         1.1877     0.2727   4.355 1.33e-05 ***
chist6         1.5071     0.2525   5.968 2.40e-09 ***
phistyes       1.4046     0.2351   5.975 2.31e-09 ***
insuranceyes   4.4462     0.5659   7.857 3.95e-15 ***
afamyes        0.5299     0.2020   2.624 0.008696 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1016.0  on 1893  degrees of freedom
AIC: 1038

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  
-2.9383  -0.4263  -0.3119  -0.2219   3.0687  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.5964     0.5863 -11.252  < 2e-16 ***
pirat          5.0623     0.9326   5.428 5.70e-08 ***
lvrat          2.1663     0.5957   3.636 0.000277 ***
chist2         0.6378     0.2349   2.715 0.006637 ** 
chist3         0.6854     0.3578   1.916 0.055385 .  
chist4         1.4107     0.3803   3.710 0.000207 ***
chist5         1.2025     0.2736   4.395 1.11e-05 ***
chist6         1.5011     0.2524   5.947 2.73e-09 ***
phistyes       1.4213     0.2358   6.027 1.67e-09 ***
insuranceyes   4.4289     0.5666   7.816 5.45e-15 ***
afamyes        0.4951     0.2030   2.439 0.014717 *  
singleyes      0.2901     0.1698   1.708 0.087626 .  
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1013.1  on 1892  degrees of freedom
AIC: 1037.1

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  
-3.0201  -0.4428  -0.3184  -0.2232   3.0774  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -5.6056     0.7376  -7.600 2.96e-14 ***
pirat          5.2335     0.9153   5.718 1.08e-08 ***
lvrat          2.2710     0.5908   3.844 0.000121 ***
chist2         0.6837     0.2329   2.936 0.003325 ** 
chist3         0.9553     0.3444   2.774 0.005541 ** 
chist4         1.3085     0.3821   3.424 0.000617 ***
chist5         1.3324     0.2654   5.020 5.17e-07 ***
chist6         1.9027     0.2363   8.051 8.24e-16 ***
insuranceyes   4.4061     0.5669   7.772 7.75e-15 ***
afamyes        0.5784     0.1966   2.941 0.003267 ** 
singleyes      0.2797     0.1674   1.671 0.094644 .  
hschoolyes    -1.0924     0.4687  -2.331 0.019770 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1042.3  on 1892  degrees of freedom
AIC: 1066.3

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  
-2.9383  -0.4263  -0.3119  -0.2219   3.0687  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.5964     0.5863 -11.252  < 2e-16 ***
pirat          5.0623     0.9326   5.428 5.70e-08 ***
lvrat          2.1663     0.5957   3.636 0.000277 ***
chist2         0.6378     0.2349   2.715 0.006637 ** 
chist3         0.6854     0.3578   1.916 0.055385 .  
chist4         1.4107     0.3803   3.710 0.000207 ***
chist5         1.2025     0.2736   4.395 1.11e-05 ***
chist6         1.5011     0.2524   5.947 2.73e-09 ***
phistyes       1.4213     0.2358   6.027 1.67e-09 ***
insuranceyes   4.4289     0.5666   7.816 5.45e-15 ***
afamyes        0.4951     0.2030   2.439 0.014717 *  
singleyes      0.2901     0.1698   1.708 0.087626 .  
---
Signif. codes:  0 ***  0.001 **  0.01 *  0.05 .  0.1    1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1013.1  on 1892  degrees of freedom
AIC: 1037.1

Number of Fisher Scoring iterations: 6
```

Hide

```
model2 <- glm(deny~pirat+lvrat+chist+phist+selfemp+insurance+afam+single+hschool,family=binomial
(link='logit'),data=train)
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  
-2.9383  -0.4263  -0.3119  -0.2219   3.0687  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -6.5964     0.5863 -11.252  < 2e-16 ***
pirat          5.0623     0.9326   5.428 5.70e-08 ***
lvrat          2.1663     0.5957   3.636 0.000277 ***
chist2         0.6378     0.2349   2.715 0.006637 ** 
chist3         0.6854     0.3578   1.916 0.055385 .  
chist4         1.4107     0.3803   3.710 0.000207 ***
chist5         1.2025     0.2736   4.395 1.11e-05 ***
chist6         1.5011     0.2524   5.947 2.73e-09 ***
phistyes       1.4213     0.2358   6.027 1.67e-09 ***
insuranceyes   4.4289     0.5666   7.816 5.45e-15 ***
afamyes        0.4951     0.2030   2.439 0.014717 *  
singleyes      0.2901     0.1698   1.708 0.087626 .  
---
Signif. codes:  0 ***  0.001 **  0.01 *  0.05 .  0.1    1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1363.1  on 1903  degrees of freedom
Residual deviance: 1013.1  on 1892  degrees of freedom
AIC: 1037.1

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
```

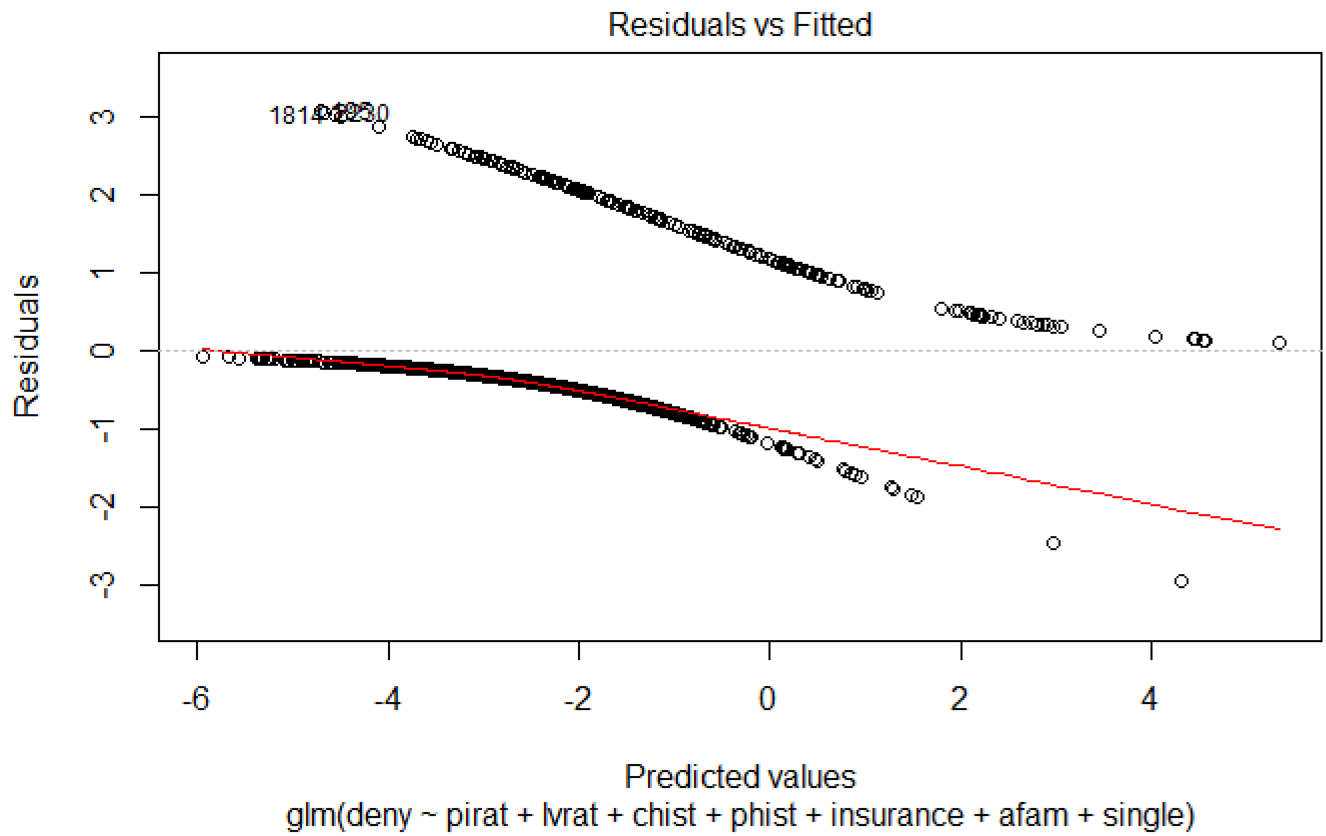
Hide

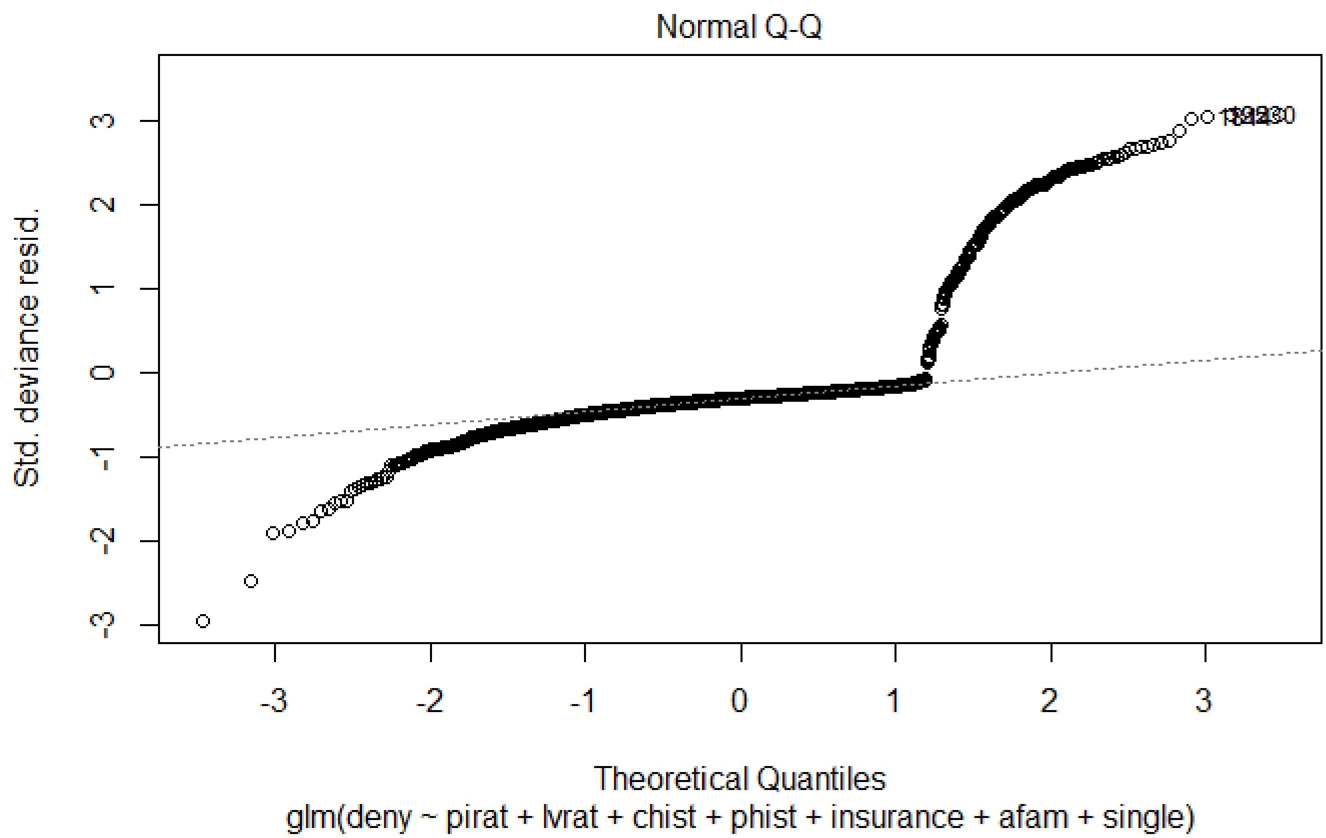
```
fm3 <- lm(I(as.numeric(deny) - 1) ~., data = dataset)
summary(fm3)
```

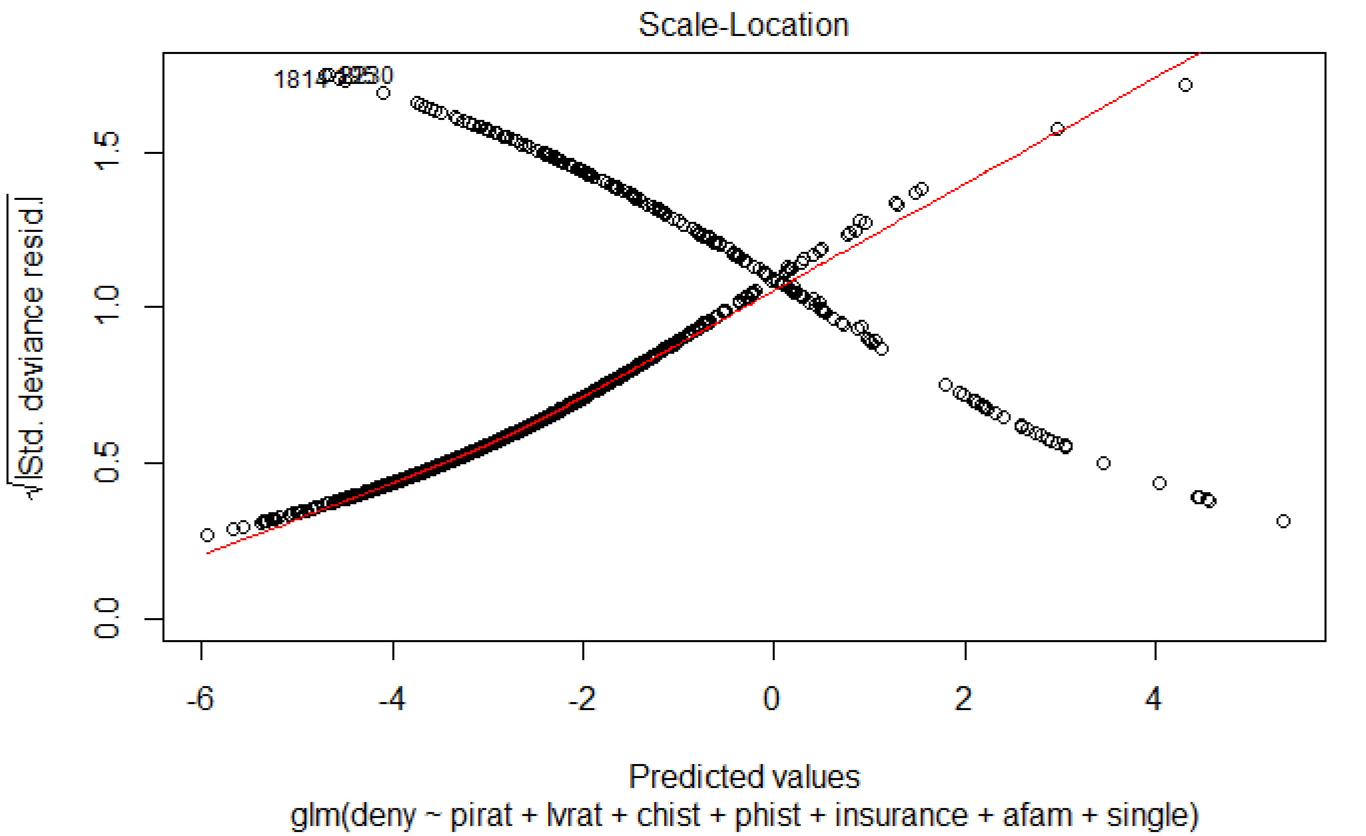
```
Call:
lm(formula = I(as.numeric(deny) - 1) ~ ., data = dataset)

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
```

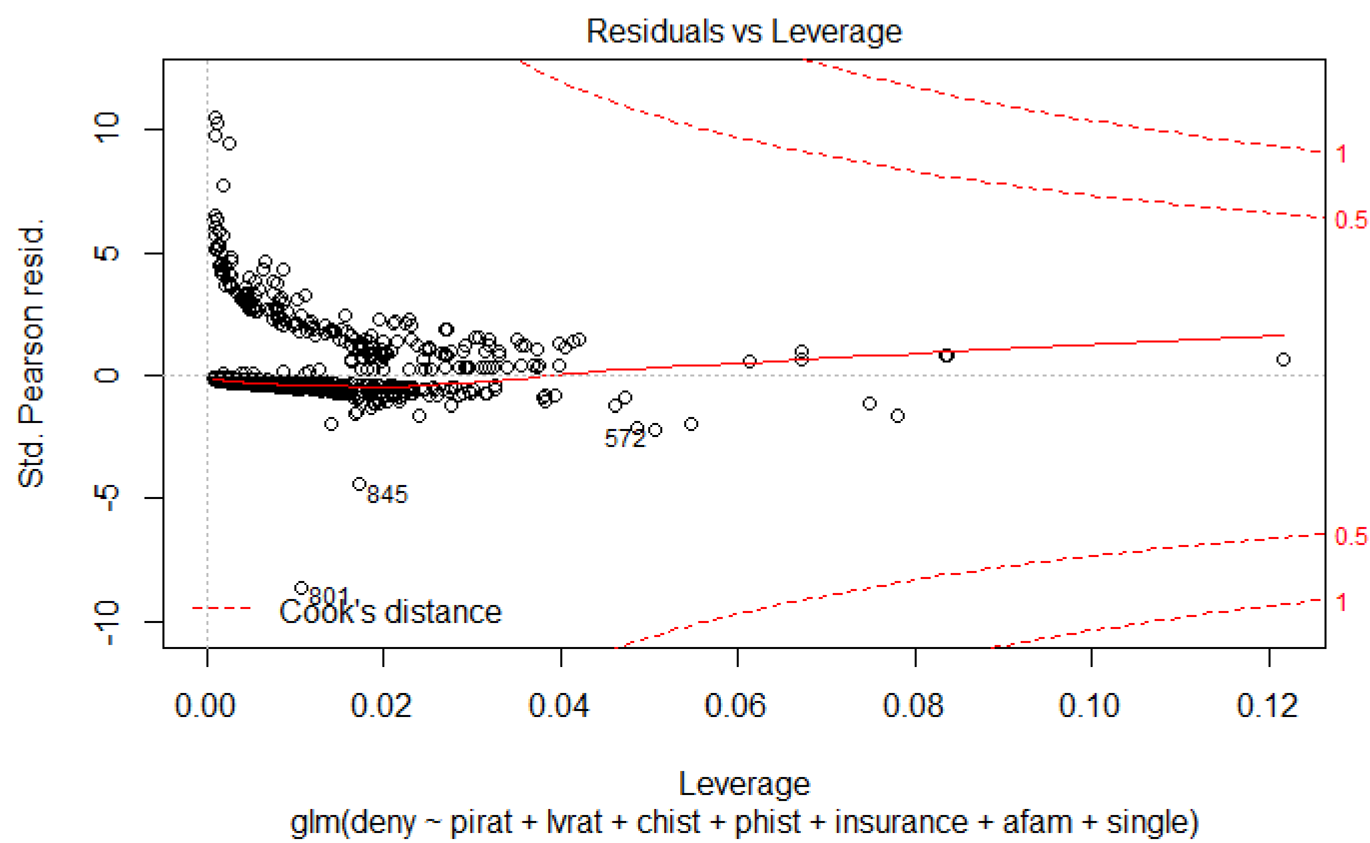
Hide

```
plot(model3)
```


Residuals vs Fitted
Residuals
-3 -2 -1 0 1 2 3
-6 -4 -2 0 2 4
Predicted values
glm(deny ~ pirat + lvrat + chist + phist + insurance + afam + single)


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





Partial F-Test

Hide

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

```
Analysis of Deviance Table

Model 1: deny ~ pirat + hirat + lvrat + chist + mhist + phist + unemp +
    selfemp + insurance + condomin + afam + single + hschool
Model 2: deny ~ pirat + lvrat + chist + phist + insurance + afam + single
  Resid. Df Resid. Dev Df Deviance
1      1884     993.29
2      1892    1013.15 -8  -19.859
```

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
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
7	no	0.350	0.290	0.7466667	1	2	no	3.9	no
8	no	0.280	0.220	0.8500000	2	2	no	1.8	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
2375	no	0.33	0.16	0.8030303	5	1	no	3.2	no
2376	no	0.31	0.25	0.8000000	1	1	no	3.2	yes
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

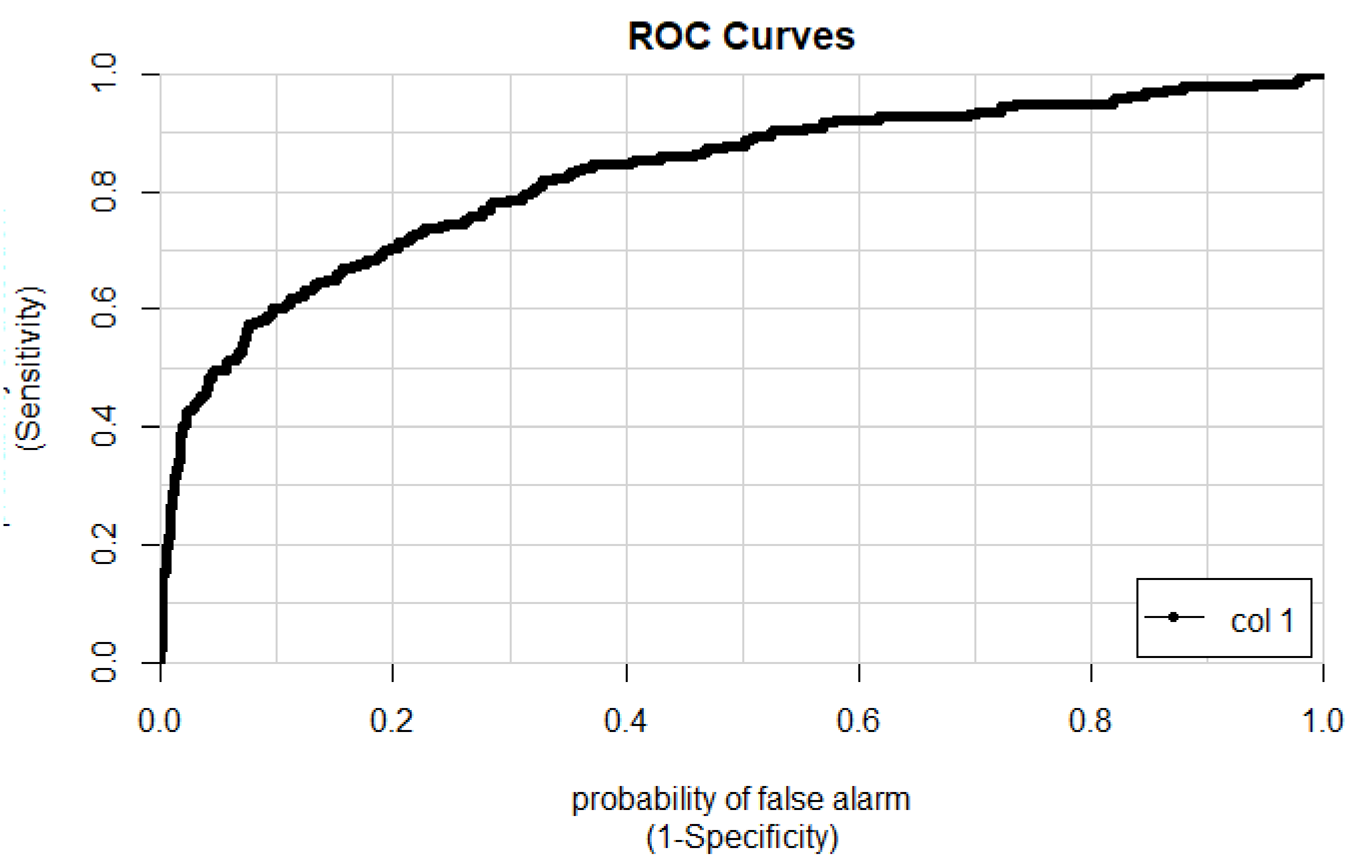
	deny <fctr>	pirat <dbl>	hirat <dbl>	lvrat <dbl>	chist <fctr>	mhist <fctr>	phist <fctr>	unemp <dbl>	selfemp <fctr>
2380	yes	0.35	0.26	0.8135593	2	2	no	4.3	no

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 0.8277181
```



Hide

```
predict1 <- ifelse(train$model3>0.95, 1, 0)
tab1 <- table(predicted = predict1, actual = train$deny)
tab1
```

```
         actual
predicted   no  yes
        0 1683  212
        1    1    8
```

Hide

```
accuracy1<-(tab1[1,1]+tab1[2,2])/(tab1[1,1]+tab1[2,2]+tab1[1,2]+tab1[2,1])
recall1<-(tab1[2,2])/(tab1[1,2]+tab1[2,2])
precision1<-(tab1[2,2])/(tab1[2,2]+tab1[2,1])
print(c("Accuracy:",accuracy1))
```

```
[1] "Accuracy:"        "0.88813025210084"
```

Hide

```
print(c("Precision:",precision1))
```

```
[1] "Precision:"        "0.888888888888889"
```

Hide

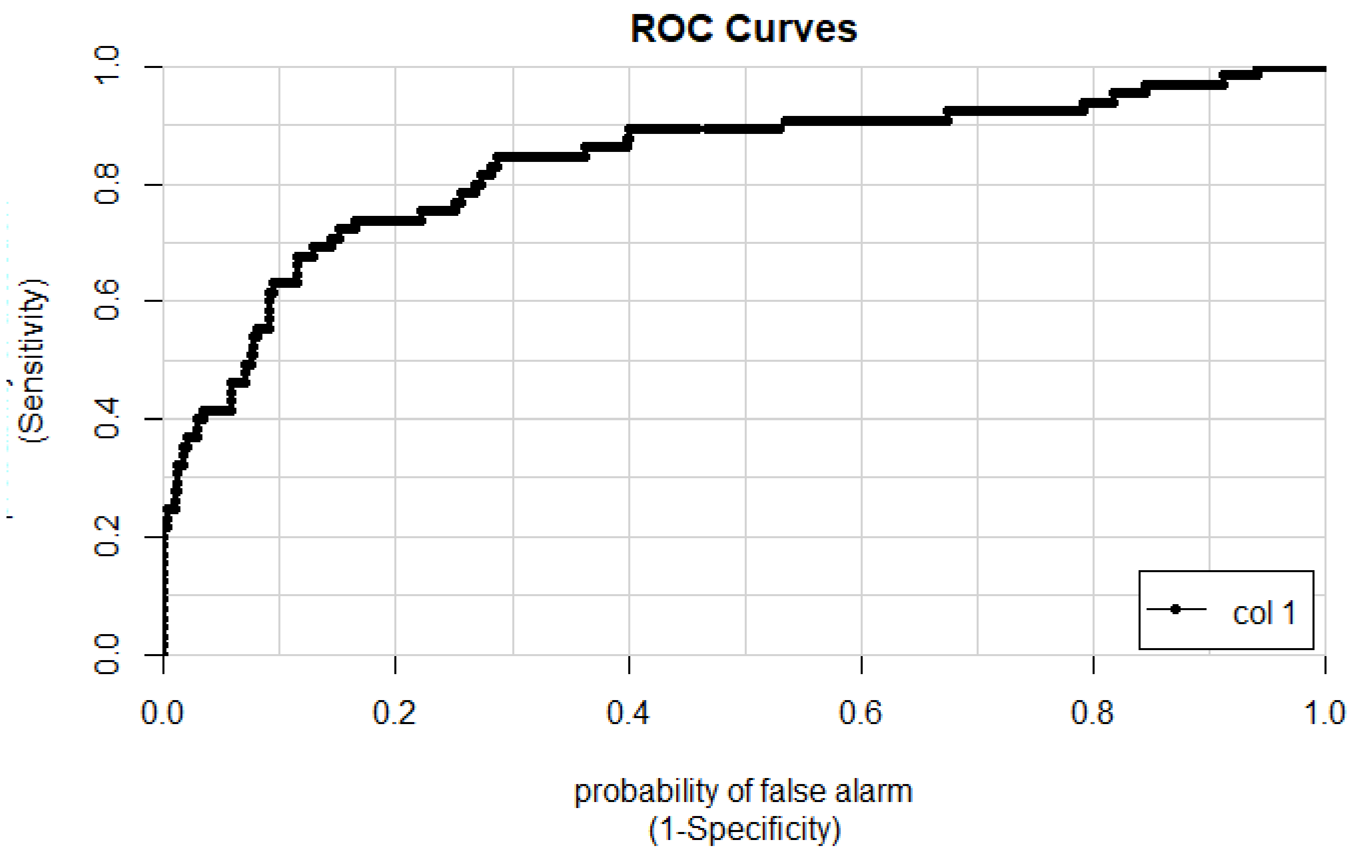
```
print(c("Recall:",recall1))
```

```
[1] "Recall:"            "0.0363636363636364"
```

Hide

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

```
                [,1]
no vs. yes 0.8348493
```



Hide

```
predict2 <- ifelse(test$model3>0.95, 1,0)
tab2 <- table(predicted = predict2, actual = test$deny)
tab2
```

```
         actual
predicted  no yes
        0 411  56
        1   0   9
```

Hide

```
accuracy<-(tab2[1,1]+tab2[2,2])/(tab2[1,1]+tab2[2,2]+tab2[1,2]+tab2[2,1])
recall<-(tab2[2,2])/(tab2[1,2]+tab2[2,2])
precision<-(tab2[2,2])/(tab2[2,2]+tab2[2,1])
print(c("Accuracy:",accuracy))
```

```
[1] "Accuracy:"         "0.882352941176471"
```

Hide

```
print(c("Precision:",precision))
```

```
[1] "Precision:" "1"
```

Hide

```
print(c("Recall:",recall))
```

```
[1] "Recall:"            "0.138461538461538"
```