Project On Logistic Regression

Aniruddha Mukherjee

Aim of the study:

When a bank receives a loan application, based on the applicant's profile the bank has to make a decision regarding whether to go ahead with the loan approval or not. Two types of risks are associated with the bank's decision:

1)If the applicant is a good credit risk, i.e. he is likely to repay the loan, then not approving the loan to the person results in a loss of business to the bank..

2)If the applicant is a bad credit risk, i.e. is not likely to repay the loan, then approving the loan to the person results in a financial loss to the bank or a hazardf to the bank workers.

It is obvious that the second risk is a greater risk, as the bank had a higher chance of not being paid back the given amount of money.

So its on the part of the bank or other lending authority to evaluate the risks associated with lending money to a customer.

This study aims at addressing this problem by using the applicant's demographic and socio-economic profiles to assess the risk of lending loan to the customer.

In business terms, we try to minimize the risk and maximize of profit for the bank. To minimize loss from the bank's perspective, the bank needs a decision rule regarding who to give approval of the loan and who not to. An applicant's demographic and socio-economic profiles are considered by loan managers before a decision is taken regarding his/her loan application.

If Y denotes the random variable that a customer is a good credit risk or a bad credit risk Then Y is dicotomous. If probability of a customer being a good credit risk is "p" then we want to guess this p based on the given data of the customer.

Description of the data:

There are 1000 observations in this dataset.The data contains data on 20 variables and the classification whether an applicant is considered a Good or a Bad credit risk for 1000 loan applicant.Here the data is a dichotomus data.

So for predicting the probability of a customer being good credit risk we use logistic regression.

We first call the data set and try to understand the catagory of the variables.

str() function can help us to know type of variables and a few sample values of each variable.

```
data=read.csv("C:\\Users\\HP\\Desktop\\gc1.csv",sep=',')
str(data)

## 'data.frame': 1000 obs. of  21 variables:
##  $ Creditability                    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Account.Balance                  : int  1 1 2 1 1 1 1 1 4 2 ...
##  $ Duration.of.Credit..month.       : int  18 9 12 12 12 10 8 6 18 24 ...
##  $ Payment.Status.of.Previous.Credit: int  4 4 2 4 4 4 4 4 4 2 ...
##  $ Purpose                          : int  2 0 9 0 0 0 0 0 3 3 ...
##  $ Credit.Amount                    : int  1049 2799 841 2122 2171 2241 3398 1361 1098 37
##  $ Value.Savings.Stocks             : int  1 1 2 1 1 1 1 1 1 3 ...
##  $ Length.of.current.employment     : int  2 3 4 3 3 2 4 2 1 1 ...
##  $ Instalment.per.cent              : int  4 2 2 3 4 1 1 2 4 1 ...
##  $ Sex...Marital.Status             : int  2 3 2 3 3 3 3 3 2 2 ...
##  $ Guarantors                       : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Duration.in.Current.address      : int  4 2 4 2 4 3 4 4 4 4 ...
##  $ Most.valuable.available.asset    : int  2 1 1 1 2 1 1 1 3 4 ...
##  $ Age..years.                      : int  21 36 23 39 38 48 39 40 65 23 ...
##  $ Concurrent.Credits               : int  3 3 3 3 1 3 3 3 3 3 ...
##  $ Type.of.apartment                : int  1 1 1 1 2 1 2 2 2 1 ...
##  $ No.of.Credits.at.this.Bank       : int  1 2 1 2 2 2 2 1 2 1 ...
##  $ Occupation                       : int  3 3 2 2 2 2 2 2 1 1 ...
##  $ No.of.dependents                 : int  1 2 1 2 1 2 1 2 1 1 ...
##  $ Telephone                        : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Foreign.Worker                   : int  1 1 1 2 2 2 2 2 1 1 ...

summary(data)

##  Creditability Account.Balance Duration.of.Credit..month.
##  Min.   :0.0   Min.   :1.000   Min.   : 4.0
##  1st Qu.:0.0   1st Qu.:1.000   1st Qu.:12.0
##  Median :1.0   Median :2.000   Median :18.0
##  Mean   :0.7   Mean   :2.577   Mean   :20.9
##  3rd Qu.:1.0   3rd Qu.:4.000   3rd Qu.:24.0
##  Max.   :1.0   Max.   :4.000   Max.   :72.0
##  Payment.Status.of.Previous.Credit    Purpose       Credit.Amount
##  Min.   :0.000                     Min.   : 0.000   Min.   :  250
##  1st Qu.:2.000                     1st Qu.: 1.000   1st Qu.: 1366
##  Median :2.000                     Median : 2.000   Median : 2320
##  Mean   :2.545                     Mean   : 2.828   Mean   : 3271
```

```
##  3rd Qu.:4.000                     3rd Qu.: 3.000   3rd Qu.: 3972
##  Max.   :4.000                     Max.   :10.000   Max.   :18424
##  Value.Savings.Stocks Length.of.current.employment Instalment.per.cent
##  Min.   :1.000        Min.   :1.000                Min.   :1.000
##  1st Qu.:1.000        1st Qu.:3.000                1st Qu.:2.000
##  Median :1.000        Median :3.000                Median :3.000
##  Mean   :2.105        Mean   :3.384                Mean   :2.973
##  3rd Qu.:3.000        3rd Qu.:5.000                3rd Qu.:4.000
##  Max.   :5.000        Max.   :5.000                Max.   :4.000
##  Sex...Marital.Status   Guarantors    Duration.in.Current.address
##  Min.   :1.000        Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000        1st Qu.:1.000   1st Qu.:2.000
##  Median :3.000        Median :1.000   Median :3.000
##  Mean   :2.682        Mean   :1.145   Mean   :2.845
##  3rd Qu.:3.000        3rd Qu.:1.000   3rd Qu.:4.000
##  Max.   :4.000        Max.   :3.000   Max.   :4.000
##  Most.valuable.available.asset  Age..years.    Concurrent.Credits
##  Min.   :1.000                 Min.   :19.00   Min.   :1.000
##  1st Qu.:1.000                 1st Qu.:27.00   1st Qu.:3.000
##  Median :2.000                 Median :33.00   Median :3.000
##  Mean   :2.358                 Mean   :35.54   Mean   :2.675
##  3rd Qu.:3.000                 3rd Qu.:42.00   3rd Qu.:3.000
##  Max.   :4.000                 Max.   :75.00   Max.   :3.000
##  Type.of.apartment No.of.Credits.at.this.Bank   Occupation
##  Min.   :1.000     Min.   :1.000             Min.   :1.000
##  1st Qu.:2.000     1st Qu.:1.000             1st Qu.:3.000
##  Median :2.000     Median :1.000             Median :3.000
##  Mean   :1.928     Mean   :1.407             Mean   :2.904
##  3rd Qu.:2.000     3rd Qu.:2.000             3rd Qu.:3.000
##  Max.   :3.000     Max.   :4.000             Max.   :4.000
##  No.of.dependents   Telephone     Foreign.Worker
##  Min.   :1.000    Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000    1st Qu.:1.000   1st Qu.:1.000
##  Median :1.000    Median :1.000   Median :1.000
##  Mean   :1.155    Mean   :1.404   Mean   :1.037
##  3rd Qu.:1.000    3rd Qu.:2.000   3rd Qu.:1.000
##  Max.   :2.000    Max.   :2.000   Max.   :2.000
```

We see that all the variables are of integer type some are continious and some arew catagorical.

And we also see the summary of the whole dataset.

Extracting the response variable:

```r
Creditability=data.frame(data$Creditability)
```

Fitting the model:

a) Logit link

```r
model=glm(Creditability~.-Creditability,data,family=binomial(link=logit))
summary(model)

##
## Call:
## glm(formula = Creditability ~ . - Creditability, family = binomial(link = logit),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5854  -0.7927   0.4512   0.7445   1.9483
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -3.994e+00  1.024e+00  -3.901 9.58e-05
## Account.Balance                   5.799e-01  7.004e-02   8.280  < 2e-16
## Duration.of.Credit..month.       -2.457e-02  8.725e-03  -2.816 0.004862
## Payment.Status.of.Previous.Credit  3.822e-01  8.740e-02   4.373 1.23e-05
## Purpose                           3.153e-02  3.009e-02   1.048 0.294697
## Credit.Amount                    -9.340e-05  4.012e-05  -2.328 0.019908
## Value.Savings.Stocks              2.391e-01  5.827e-02   4.104 4.07e-05
## Length.of.current.employment      1.517e-01  7.118e-02   2.132 0.033027
## Instalment.per.cent              -2.983e-01  8.276e-02  -3.605 0.000312
## Sex...Marital.Status              2.574e-01  1.157e-01   2.224 0.026131
## Guarantors                        3.473e-01  1.777e-01   1.954 0.050681
## Duration.in.Current.address      -1.411e-02  7.742e-02  -0.182 0.855335
## Most.valuable.available.asset    -1.828e-01  9.101e-02  -2.009 0.044521
## Age..years.                       8.917e-03  8.206e-03   1.087 0.277218
## Concurrent.Credits                2.419e-01  1.111e-01   2.178 0.029420
## Type.of.apartment                 2.931e-01  1.677e-01   1.748 0.080527
## No.of.Credits.at.this.Bank       -2.436e-01  1.610e-01  -1.513 0.130257
## Occupation                        1.889e-02  1.367e-01   0.138 0.890081
## No.of.dependents                 -1.708e-01  2.319e-01  -0.736 0.461567
## Telephone                         2.947e-01  1.880e-01   1.567 0.117024
## Foreign.Worker                    1.158e+00  6.078e-01   1.906 0.056680
##
## (Intercept)                      ***
```

```
## Account.Balance                   ***
## Duration.of.Credit..month.        **
## Payment.Status.of.Previous.Credit ***
## Purpose
## Credit.Amount                     *
## Value.Savings.Stocks              ***
## Length.of.current.employment      *
## Instalment.per.cent               ***
## Sex...Marital.Status              *
## Guarantors                        .
## Duration.in.Current.address
## Most.valuable.available.asset     *
## Age..years.
## Concurrent.Credits                *
## Type.of.apartment                 .
## No.of.Credits.at.this.Bank
## Occupation
## No.of.dependents
## Telephone
## Foreign.Worker                    .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.73  on 999  degrees of freedom
## Residual deviance:  956.56  on 979  degrees of freedom
## AIC: 998.56
##
## Number of Fisher Scoring iterations: 5
```

b)probit Link

```
model1=glm(Creditability~.-Creditability,data,family=binomial(link=probit))
summary(model1)
```

```
##
## Call:
## glm(formula = Creditability ~ . - Creditability, family = binomial(link = probit),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6658  -0.8066   0.4497   0.7660   1.9444
```

```
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -2.200e+00  5.785e-01  -3.804 0.000142
## Account.Balance                    3.424e-01  3.980e-02   8.603  < 2e-16
## Duration.of.Credit..month.        -1.397e-02  5.139e-03  -2.719 0.006556
## Payment.Status.of.Previous.Credit  2.220e-01  5.044e-02   4.402 1.07e-05
## Purpose                            1.728e-02  1.745e-02   0.990 0.321967
## Credit.Amount                     -5.532e-05  2.363e-05  -2.341 0.019209
## Value.Savings.Stocks               1.312e-01  3.290e-02   3.988 6.67e-05
## Length.of.current.employment       8.755e-02  4.148e-02   2.111 0.034799
## Instalment.per.cent               -1.748e-01  4.801e-02  -3.641 0.000272
## Sex...Marital.Status               1.477e-01  6.734e-02   2.193 0.028308
## Guarantors                         1.804e-01  1.020e-01   1.768 0.077000
## Duration.in.Current.address       -7.643e-03  4.520e-02  -0.169 0.865724
## Most.valuable.available.asset     -1.102e-01  5.266e-02  -2.093 0.036321
## Age..years.                        5.479e-03  4.763e-03   1.150 0.250029
## Concurrent.Credits                 1.381e-01  6.508e-02   2.122 0.033829
## Type.of.apartment                  1.629e-01  9.818e-02   1.659 0.097084
## No.of.Credits.at.this.Bank        -1.460e-01  9.306e-02  -1.569 0.116662
## Occupation                         1.015e-02  8.015e-02   0.127 0.899235
## No.of.dependents                  -1.133e-01  1.346e-01  -0.842 0.400031
## Telephone                          1.692e-01  1.084e-01   1.560 0.118670
## Foreign.Worker                     6.389e-01  3.272e-01   1.952 0.050887
##
## (Intercept)                       ***
## Account.Balance                   ***
## Duration.of.Credit..month.        **
## Payment.Status.of.Previous.Credit ***
## Purpose
## Credit.Amount                     *
## Value.Savings.Stocks              ***
## Length.of.current.employment      *
## Instalment.per.cent               ***
## Sex...Marital.Status              *
## Guarantors                        .
## Duration.in.Current.address
## Most.valuable.available.asset     *
## Age..years.
## Concurrent.Credits                *
## Type.of.apartment                 .
## No.of.Credits.at.this.Bank
## Occupation
## No.of.dependents
## Telephone
```

```
## Foreign.Worker                    .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.73  on 999  degrees of freedom
## Residual deviance:  957.74  on 979  degrees of freedom
## AIC: 999.74
##
## Number of Fisher Scoring iterations: 5
```

c) complemantary log log link.

```
model2=glm(Creditability~.-Creditability,data,family=binomial(link=cloglog))
summary(model2)
```

```
##
## Call:
## glm(formula = Creditability ~ . - Creditability, family = binomial(link = cloglog),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8413  -0.8979   0.4421   0.8042   1.8028
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -2.222e+00  5.470e-01  -4.062 4.87e-05
## Account.Balance                  3.323e-01  3.735e-02   8.895  < 2e-16
## Duration.of.Credit..month.      -1.201e-02  5.290e-03  -2.271 0.023137
## Payment.Status.of.Previous.Credit  2.114e-01  5.072e-02   4.169 3.06e-05
## Purpose                          1.076e-02  1.698e-02   0.634 0.526283
## Credit.Amount                   -5.642e-05  2.476e-05  -2.279 0.022656
## Value.Savings.Stocks             1.024e-01  2.980e-02   3.435 0.000593
## Length.of.current.employment     8.946e-02  4.022e-02   2.224 0.026141
## Instalment.per.cent             -1.585e-01  4.642e-02  -3.414 0.000641
## Sex...Marital.Status             1.400e-01  6.481e-02   2.160 0.030788
## Guarantors                       1.142e-01  9.635e-02   1.185 0.236080
## Duration.in.Current.address      1.923e-03  4.415e-02   0.044 0.965256
## Most.valuable.available.asset   -1.064e-01  5.033e-02  -2.115 0.034430
## Age..years.                      5.712e-03  4.588e-03   1.245 0.213213
## Concurrent.Credits               1.279e-01  6.570e-02   1.947 0.051499
## Type.of.apartment                1.175e-01  9.885e-02   1.188 0.234663
```

```
## No.of.Credits.at.this.Bank        -1.477e-01  9.200e-02  -1.605 0.108446
## Occupation                        -1.869e-02  7.909e-02  -0.236 0.813219
## No.of.dependents                  -1.324e-01  1.295e-01  -1.022 0.306785
## Telephone                          1.772e-01  1.030e-01   1.721 0.085273
## Foreign.Worker                     5.524e-01  2.770e-01   1.994 0.046106
##
## (Intercept)                       ***
## Account.Balance                   ***
## Duration.of.Credit..month.        *
## Payment.Status.of.Previous.Credit ***
## Purpose
## Credit.Amount                     *
## Value.Savings.Stocks              ***
## Length.of.current.employment      *
## Instalment.per.cent               ***
## Sex...Marital.Status              *
## Guarantors
## Duration.in.Current.address
## Most.valuable.available.asset     *
## Age..years.
## Concurrent.Credits                .
## Type.of.apartment
## No.of.Credits.at.this.Bank
## Occupation
## No.of.dependents
## Telephone                         .
## Foreign.Worker                    *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.7  on 999  degrees of freedom
## Residual deviance:  964.8  on 979  degrees of freedom
## AIC: 1006.8
##
## Number of Fisher Scoring iterations: 6
```

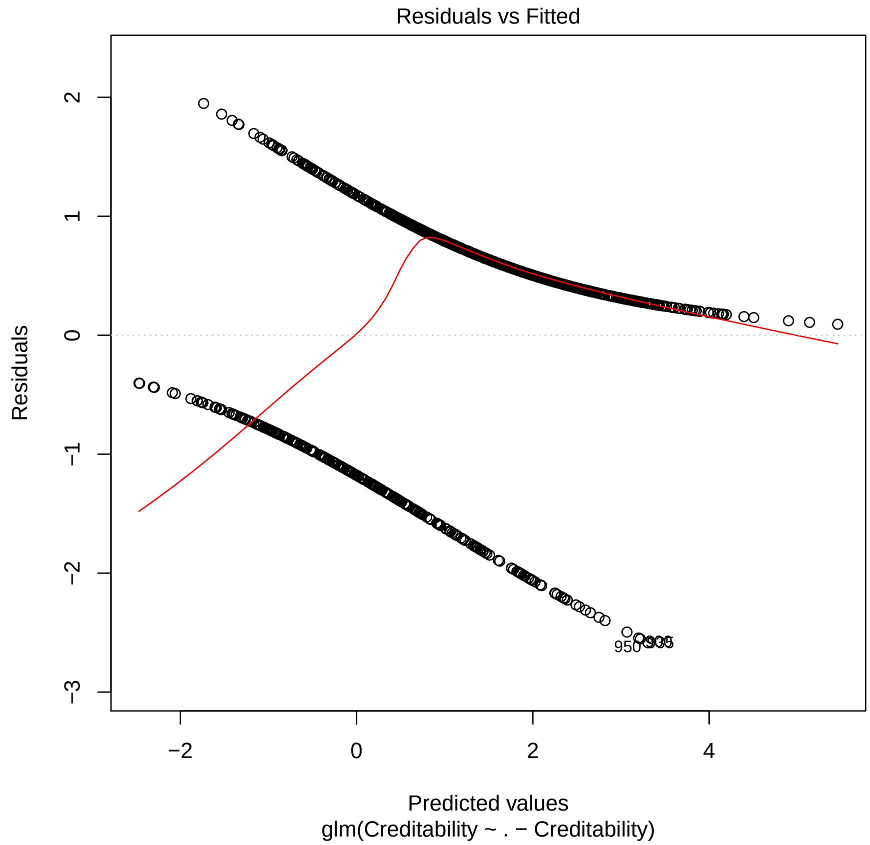
We see that the residual deviance of the logit limk is the lowest so we say that logit link is the best link here.

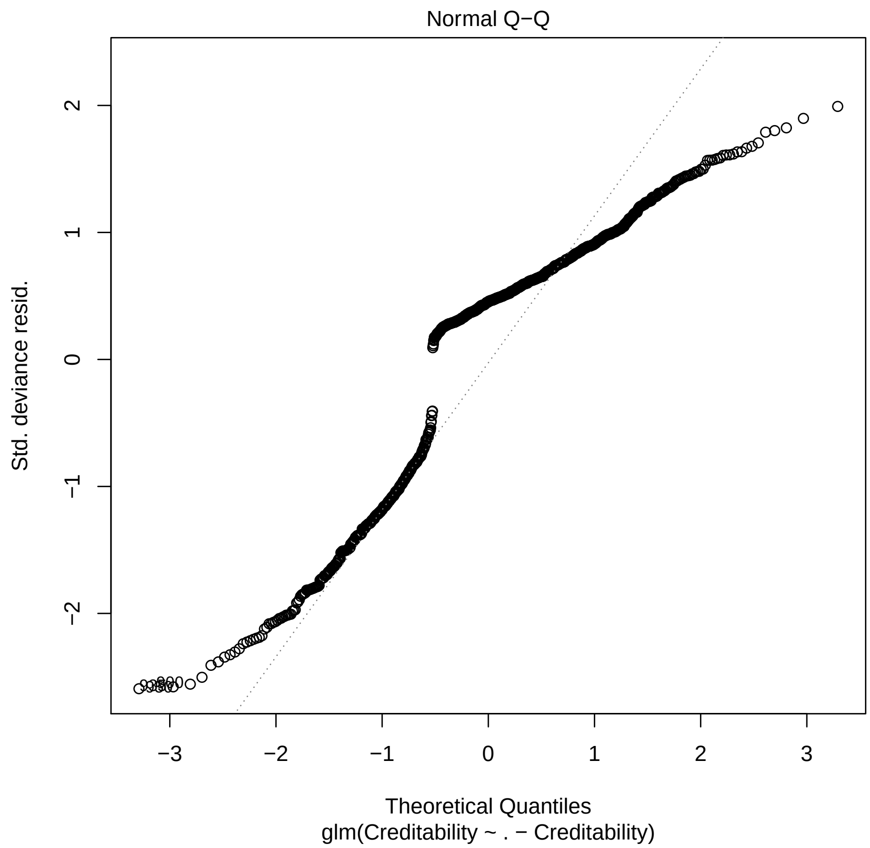
and also the p vaule for the coefficient of Foreign workers is greter than 0.05 so that variable is not important to explain the good or bad credit risk.

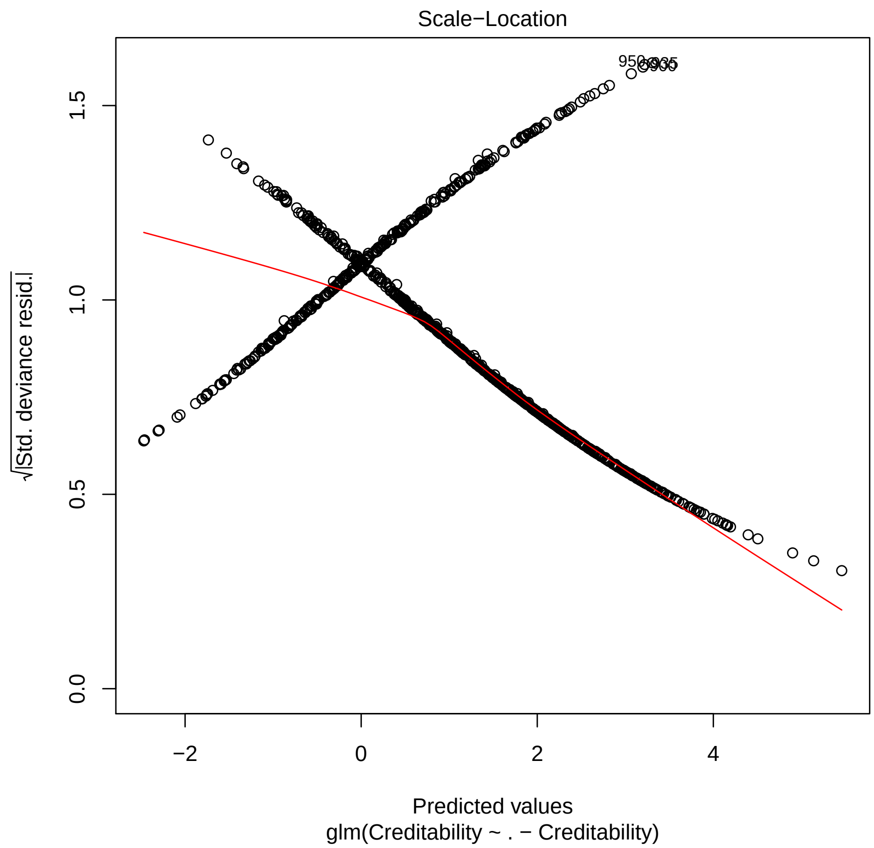
Graphical view:

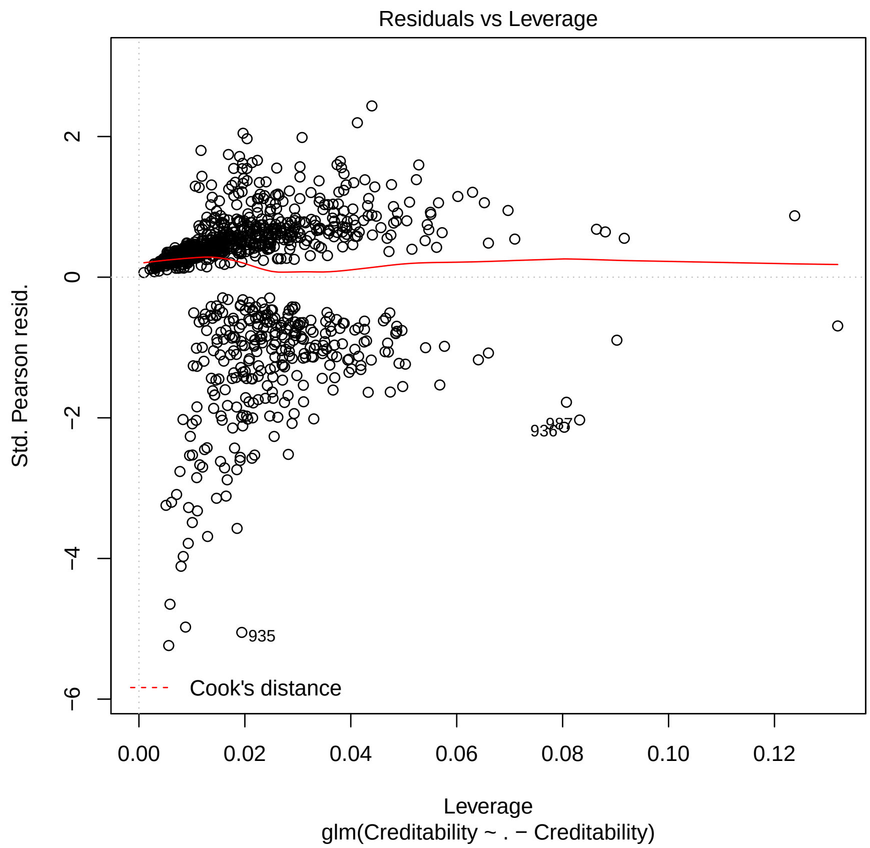
a) Logit model

```
plot(model)
```



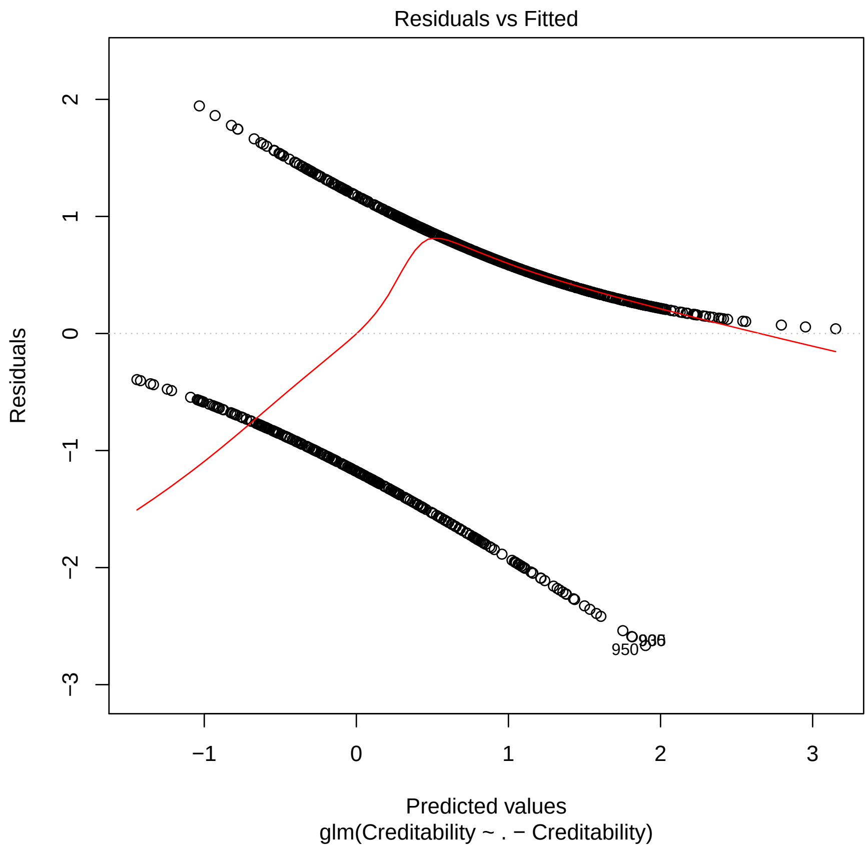

Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(Creditability ~ . − Creditability)
−3
−2
−1
0
1
2
3

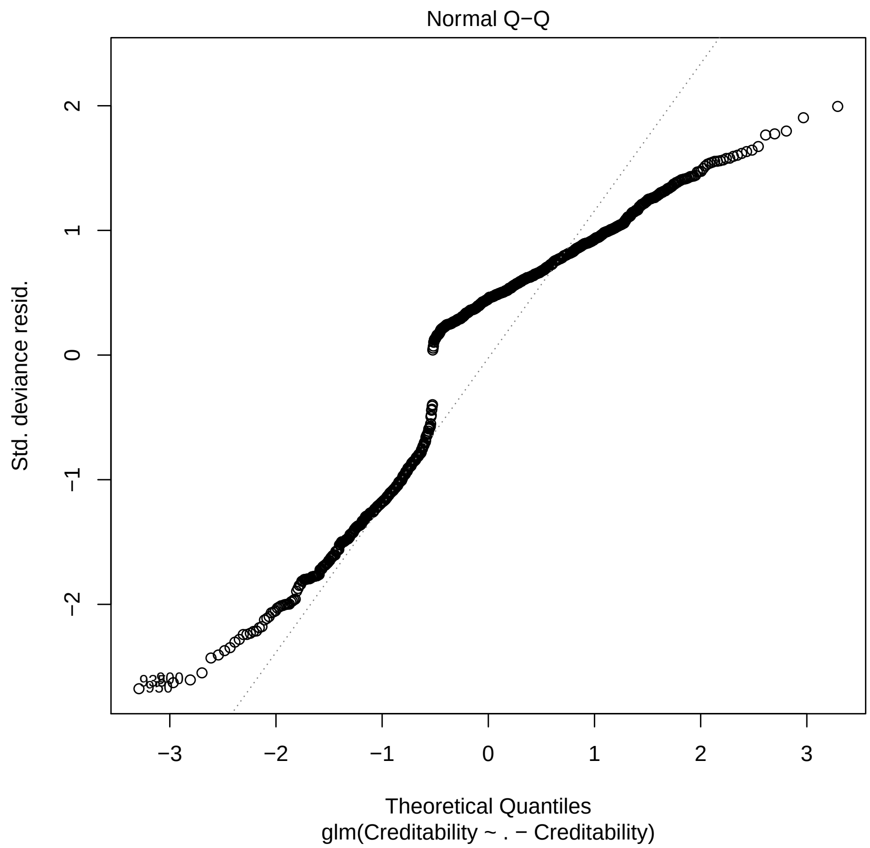

Scale−Location
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
−2
0
2
4
Predicted values
glm(Creditability ~ . − Creditability)

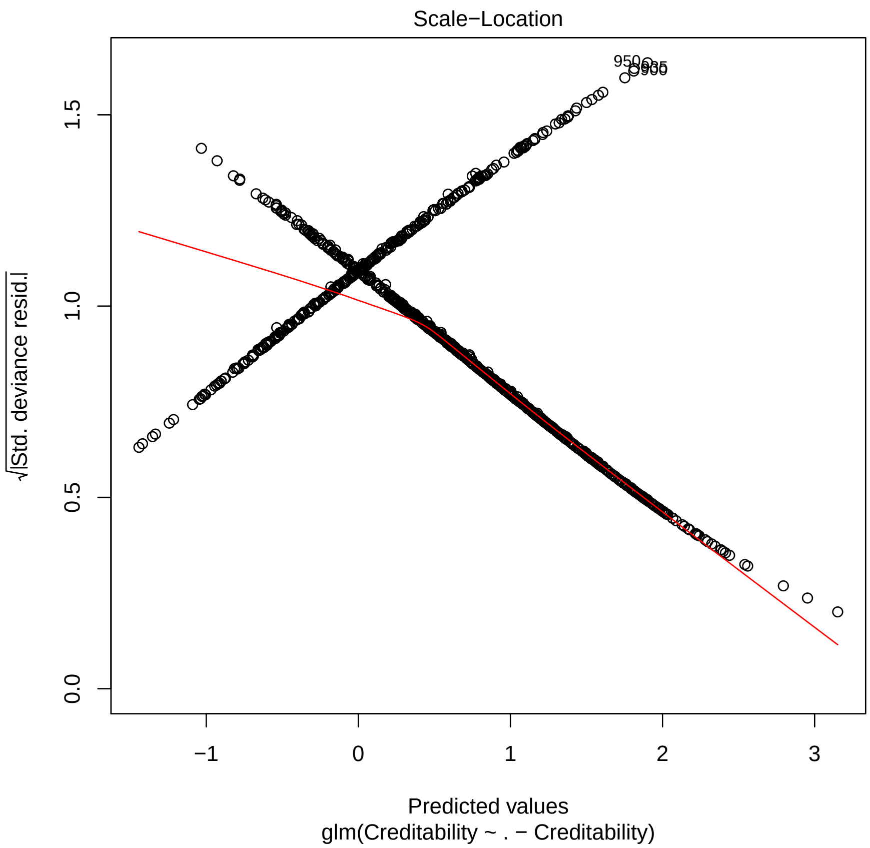


b)probit model

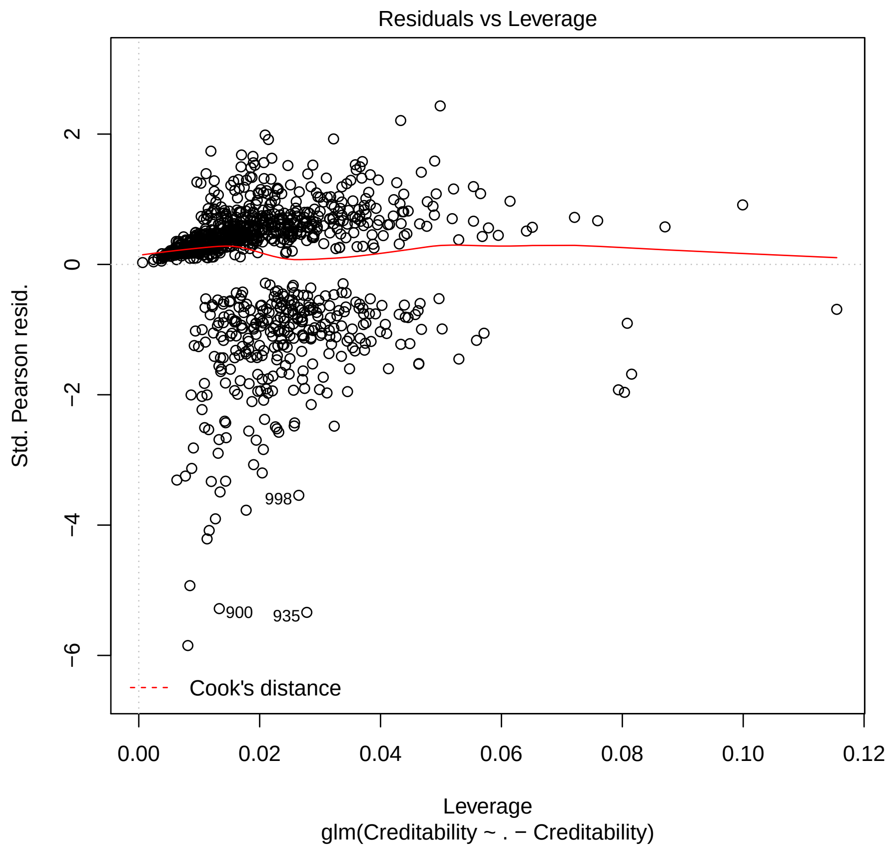
```
plot(model1)
```


Residuals vs Fitted
Residuals
2
1
0
−1
−2
−3
−1
0
1
2
3
950
936
Predicted values
glm(Creditability ~ . − Creditability)


Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(Creditability ~ . − Creditability)

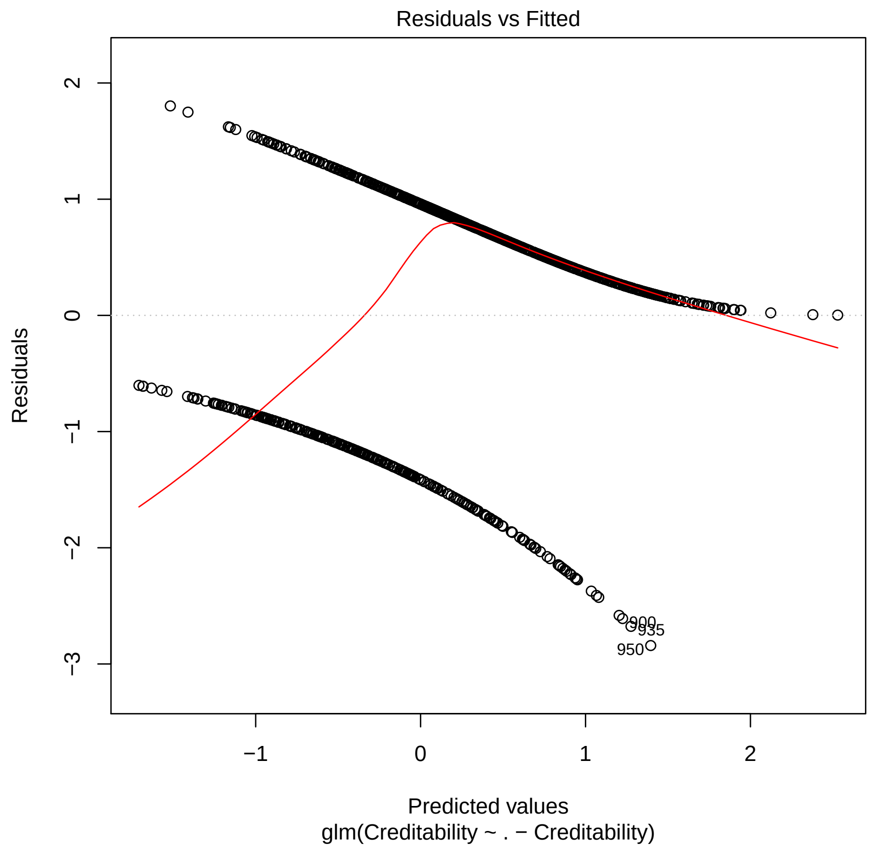


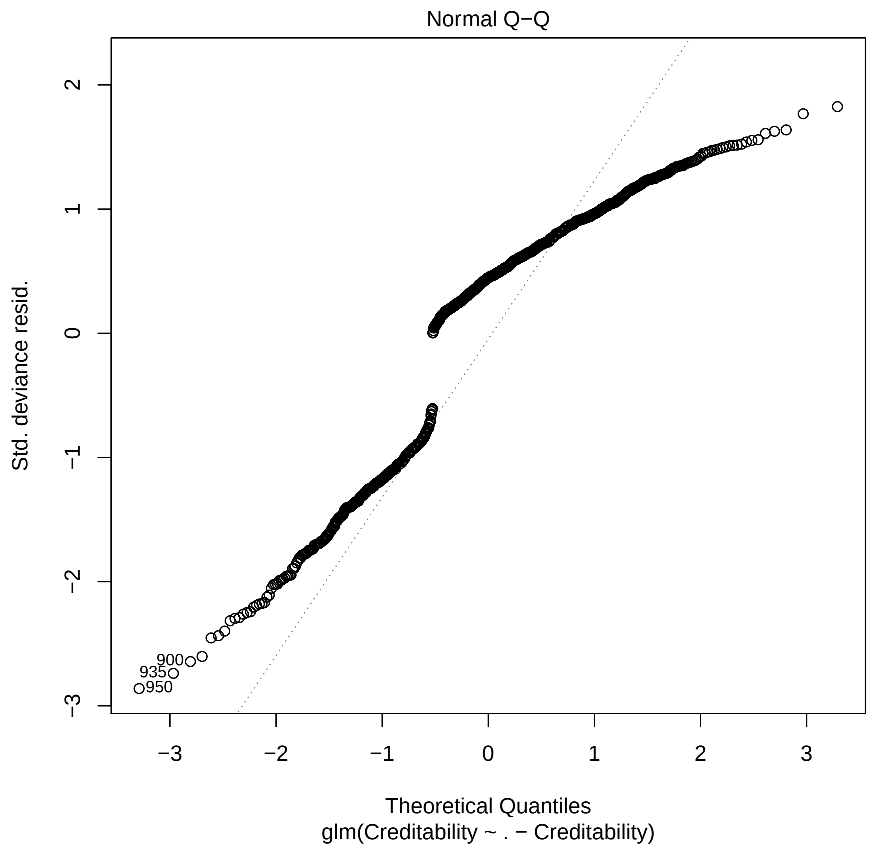
Scale−Location
950
935
900
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
−1
0
1
2
3
Predicted values
glm(Creditability ~ . − Creditability)



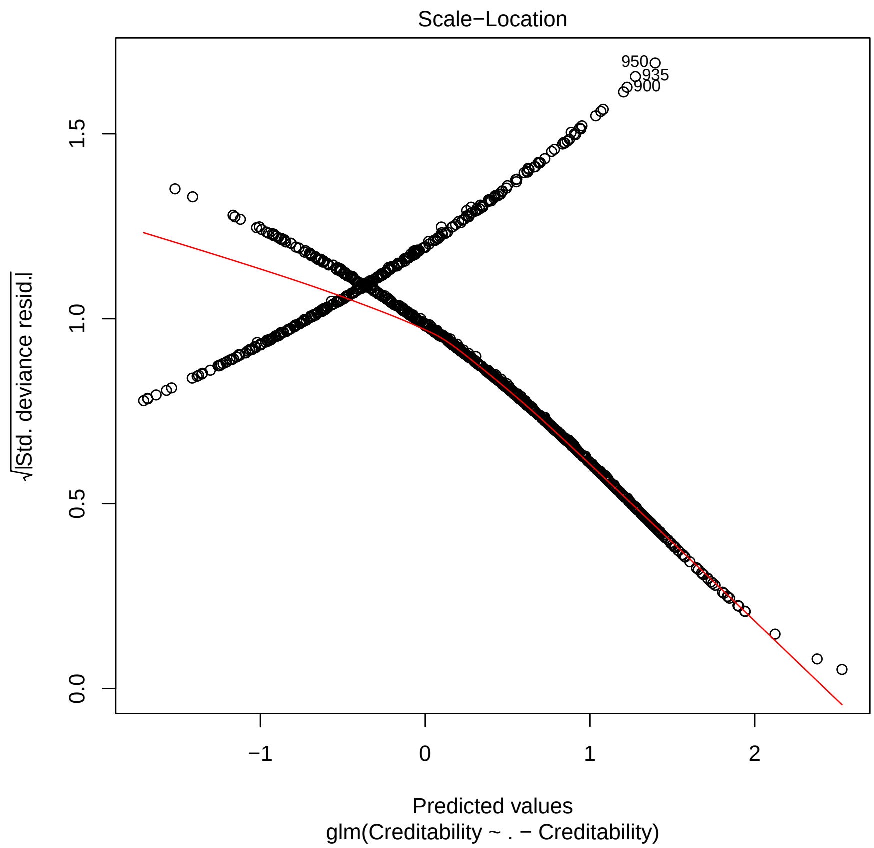
c) complemetary log log model

```
plot(model2)
```

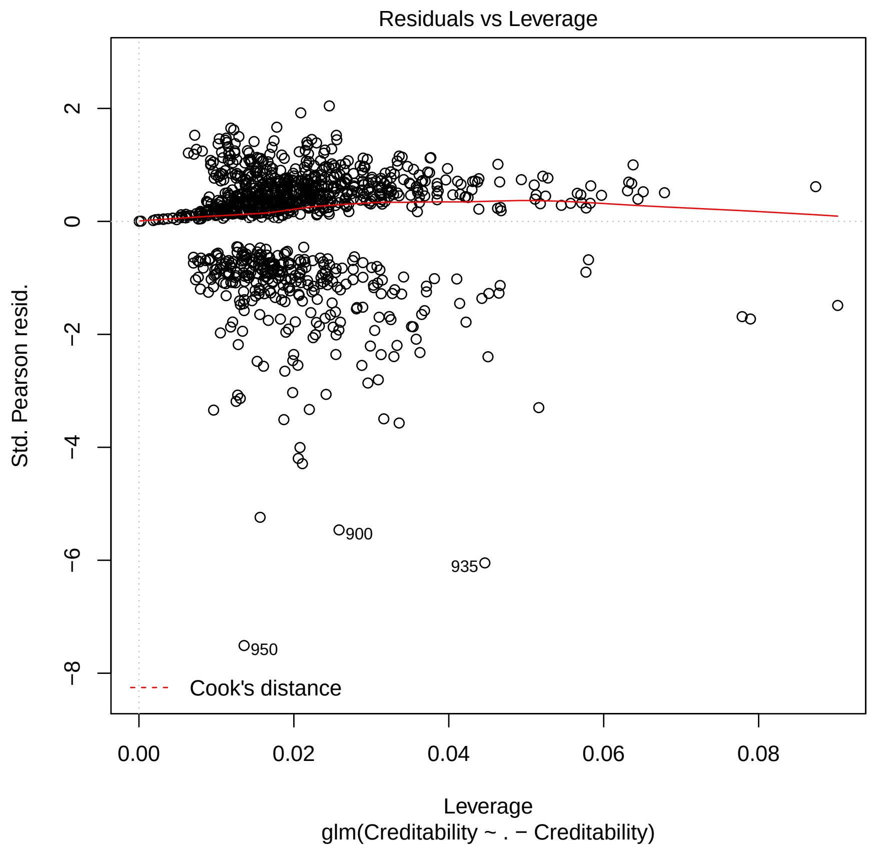

Residuals vs Fitted
Residuals
2
1
0
−1
−2
−3
−1
0
1
2
Predicted values
glm(Creditability ~ . − Creditability)
900
935
950



Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
glm(Creditability ~ . − Creditability)
900
935
950



Scale−Location
950
935
900
√|Std. deviance resid.|
0.0
0.5
1.0
1.5
−1
0
1
2
Predicted values
glm(Creditability ~ . − Creditability)



Variable selection:

```
#stepwise selection
library(leaps)

## Warning:  package 'leaps' was built under R version 3.4.4

leaps=regsubsets(Creditability~.,data=data,nbest=13,nvmax=13)
null=glm(Creditability~.,data,family=binomial)
full=glm(Creditability~.-Creditability,data,family=binomial)
step(null,scope=list(lower=null,upper=full),direction="both")

## Start:  AIC=998.56
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##     Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
```

```
##     Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##     Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##     Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##     Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##     No.of.dependents + Telephone + Foreign.Worker
##
## Call:  glm(formula = Creditability ~ Account.Balance + Duration.of.Credit..month. +
##     Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##     Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##     Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##     Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##     Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##     No.of.dependents + Telephone + Foreign.Worker, family = binomial,
##     data = data)
##
## Coefficients:
##                       (Intercept)                    Account.Balance
##                        -3.9940768                          0.5799270
##        Duration.of.Credit..month.  Payment.Status.of.Previous.Credit
##                        -0.0245701                          0.3821907
##                           Purpose                      Credit.Amount
##                         0.0315277                         -0.0000934
##              Value.Savings.Stocks       Length.of.current.employment
##                         0.2391122                          0.1517308
##               Instalment.per.cent               Sex...Marital.Status
##                        -0.2983367                          0.2573791
##                        Guarantors        Duration.in.Current.address
##                         0.3472739                         -0.0141141
##     Most.valuable.available.asset                        Age..years.
##                        -0.1828445                          0.0089167
##                Concurrent.Credits                  Type.of.apartment
##                         0.2418915                          0.2930602
##        No.of.Credits.at.this.Bank                         Occupation
##                        -0.2435882                          0.0188903
##                  No.of.dependents                          Telephone
##                        -0.1707594                          0.2946784
##                    Foreign.Worker
##                         1.1583058
##
## Degrees of Freedom: 999 Total (i.e. Null);  979 Residual
## Null Deviance:     1222
## Residual Deviance: 956.6  AIC: 998.6
```

```
#Forward selection
step(null,scope=list(lower=null,upper=full),direction="forward")
```

```
## Start:  AIC=998.56
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##     Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##     Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##     Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##     Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##     Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##     No.of.dependents + Telephone + Foreign.Worker
##
## Call:  glm(formula = Creditability ~ Account.Balance + Duration.of.Credit..month. +
##     Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##     Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##     Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##     Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##     Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##     No.of.dependents + Telephone + Foreign.Worker, family = binomial,
##     data = data)
##
## Coefficients:
##                       (Intercept)                    Account.Balance
##                        -3.9940768                          0.5799270
##        Duration.of.Credit..month.  Payment.Status.of.Previous.Credit
##                        -0.0245701                          0.3821907
##                           Purpose                      Credit.Amount
##                         0.0315277                         -0.0000934
##              Value.Savings.Stocks       Length.of.current.employment
##                         0.2391122                          0.1517308
##               Instalment.per.cent               Sex...Marital.Status
##                        -0.2983367                          0.2573791
##                        Guarantors        Duration.in.Current.address
##                         0.3472739                         -0.0141141
##     Most.valuable.available.asset                        Age..years.
##                        -0.1828445                          0.0089167
##                Concurrent.Credits                  Type.of.apartment
##                         0.2418915                          0.2930602
##        No.of.Credits.at.this.Bank                         Occupation
##                        -0.2435882                          0.0188903
##                  No.of.dependents                          Telephone
##                        -0.1707594                          0.2946784
##                    Foreign.Worker
##                         1.1583058
##
## Degrees of Freedom: 999 Total (i.e. Null);  979 Residual
## Null Deviance:     1222
## Residual Deviance: 956.6  AIC: 998.6
```

comment: Here we see that the responsible variables that we are working upon are suitable and appropriate.There is no need of further subtraction or addition of Responsible Variables.

Visualisation of the correlation matrix:

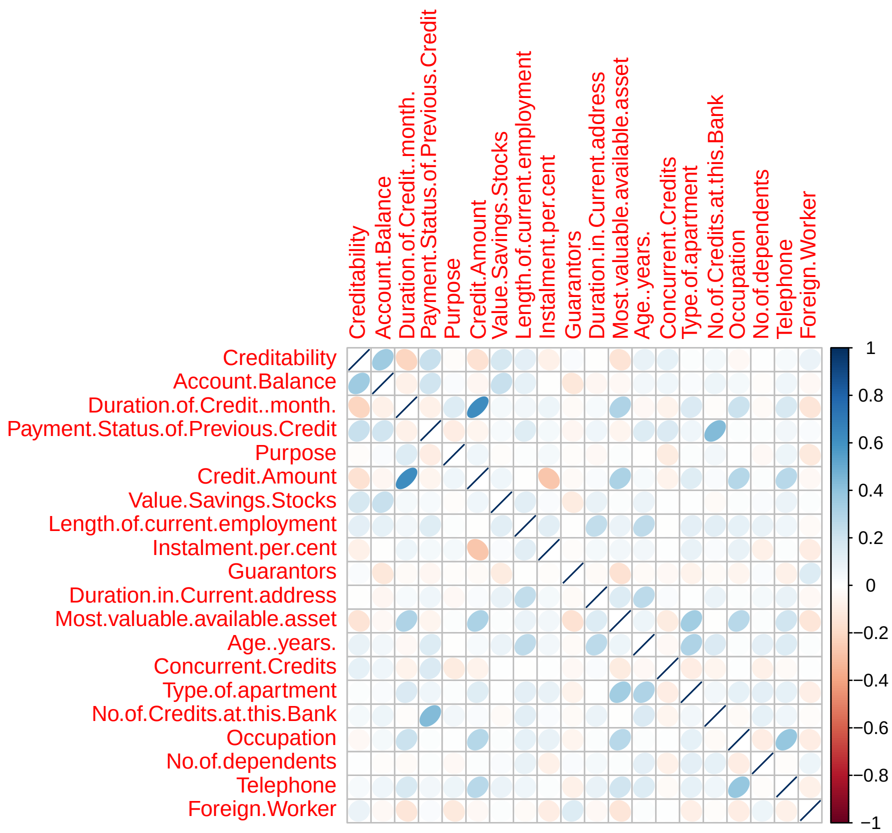
```
install.packages("corrplot")

## Installing package into 'C:/Users/HP/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"):  trying to use CRAN without
setting a mirror

library(corrplot)

## Warning:  package 'corrplot' was built under R version 3.4.4
## corrplot 0.84 loaded

corrplot(cor(data.matrix(data[,-c(10)])), method = "ellipse")
```



Here in the plot we see that intensty of the non diagonal entries are really low indicating that there is no correlation between any of the variables so mear multicolliniarity is not present in the data set...

Conclusion:

After analysing the data set we come to the conclusion that logistic regreesion with logit model is appropriate to explain the variability and presicting "P". Also after fitting from the graph of residual vs fit we see the random pattern confirming a good fit.

And also from the variable selection from forward or both side selection we could not reduce any variable. So we say that set of variables selected by the bank is really good and sufficient set of variables.

Depending on the formula developed if the bank lends money to the persons will run on profit.