HCC: Predictive Model

Ph. Nesma Magdy Ahmed

22/04/2022

Data Cleaning:

```
data <- read.csv(choose.files(), header = T, stringsAsFactors = T)
attach(data)

data[sapply(data[ ,1:23], is.numeric)] <- lapply(data[sapply(data[ ,1:23], is.numeric)],
                                    as.factor)
dt <- read.csv(choose.files(), header = T, stringsAsFactors = T)
attach(dt)
```

Demographic Characheristics:

```
data %>% tbl_summary(by=Class) %>%
  bold_labels() %>%
  add_labelled_class()%>%
   modify_footnote(
     all_stat_cols() ~ "Median (IQR) or Frequency (%)"
  ) %>%
  add_p() %>%
  modify_header(label ~ "**Participants Characteristics**") %>%
  modify_caption("**Table 1. Patients' Characteristics (n = 204)**") %>%
  bold_labels() %>%
  bold_p()
```

Table 1: **Table 1. Patients' Characteristics (n = 204)**

Participants Characteristics	**Died**, N = 102	**Lived**, N = 102	p-value
Gender			>0.9
Female	21 (21%)	21 (21%)	
Male	81 (79%)	81 (79%)	
Symptoms			**<0.001**
No Symptoms	18 (18%)	45 (44%)	
Symptoms	84 (82%)	57 (56%)	
Alcohol			0.9
Not Alcoholic	29 (28%)	28 (27%)	
Alcoholic	73 (72%)	74 (73%)	
HBsAg			0.5

Participants Characteristics	**Died**, N = 102	**Lived**, N = 102	**p-value**
No Hepatitis B Surface Antigen	94 (92%)	91 (89%)	
Hepatitis B Surface Antigen	8 (7.8%)	11 (11%)	
HBeAg			0.2
No Hepatitis B e Antigen	99 (97%)	102 (100%)	
Hepatitis B e Antigen	3 (2.9%)	0 (0%)	
HBcAb			0.4
No Hepatitis B Core Antibody	78 (76%)	73 (72%)	
Hepatitis B Core Antibody	24 (24%)	29 (28%)	
HCVAb			0.3
No Hepatitis C Virus Antibody	78 (76%)	84 (82%)	
Hepatitis C Virus Antibody	24 (24%)	18 (18%)	
Cirrhosis			0.5
No Cirrhosis	12 (12%)	9 (8.8%)	
Cirrhosis	90 (88%)	93 (91%)	
Endemic			**0.017**
Not living in Endemic Countries	100 (98%)	92 (90%)	
Living in Endemic Countries	2 (2.0%)	10 (9.8%)	
Smoking			0.091
Not Smoker	62 (61%)	50 (49%)	
Smoker	40 (39%)	52 (51%)	
Diabetes			0.058
Not Diabetic	59 (58%)	72 (71%)	
Diabetic	43 (42%)	30 (29%)	
Obesity			0.4
Normal Weight	93 (91%)	89 (87%)	
Obese	9 (8.8%)	13 (13%)	
Hemochro			0.8
No Hemochromatosis	96 (94%)	97 (95%)	
Hemochromatosis	6 (5.9%)	5 (4.9%)	
AHT			0.7
No Arterial Hypertension	66 (65%)	63 (62%)	
Arterial Hypertension	36 (35%)	39 (38%)	
CRI			0.071
No Chronic Renal Insufficiency	83 (81%)	92 (90%)	
Chronic Renal Insufficiency	19 (19%)	10 (9.8%)	
HIV			>0.9
No Human Immunodeficiency Virus (HIV)	101 (99%)	100 (98%)	
Human Immunodeficiency Virus (HIV)	1 (1.0%)	2 (2.0%)	
NASH			0.2
No Nonalcoholic Steatohepatitis	99 (97%)	95 (93%)	
Nonalcoholic Steatohepatitis	3 (2.9%)	7 (6.9%)	
Varices			0.9
No Esophageal Varices	37 (36%)	38 (37%)	
Esophageal Varices	65 (64%)	64 (63%)	
Spleno			0.5
No Splenomegaly	50 (49%)	45 (44%)	
Splenomegaly	52 (51%)	57 (56%)	
PHT			0.9
No Portal Hypertension	33 (32%)	32 (31%)	
Portal Hypertension	69 (68%)	70 (69%)	
PVT			**0.011**
No Portal Vein Thrombosis	72 (71%)	87 (85%)	

Participants Characteristics	Died, N = 102	Lived, N = 102	p-value
Portal Vein Thrombosis	30 (29%)	15 (15%)	
Metastasis			**<0.001**
No Liver Metastasis	65 (64%)	88 (86%)	
Liver Metastasis	37 (36%)	14 (14%)	
Hallmark			0.2
No Radiological Hallmark	39 (38%)	31 (30%)	
Radiological Hallmark	63 (62%)	71 (70%)	
Age	34 (25, 41)	28 (20, 37)	**0.005**
Grams_day	21 (7, 30)	30 (1, 30)	0.5
Packs_year	3 (0, 30)	0 (0, 40)	0.9
PS			**<0.001**
0	25 (25%)	63 (62%)	
1	28 (27%)	16 (16%)	
2	20 (20%)	17 (17%)	
3	23 (23%)	6 (5.9%)	
4	6 (5.9%)	0 (0%)	
Encephalopathy			**0.017**
1	82 (80%)	91 (89%)	
2	13 (13%)	11 (11%)	
3	7 (6.9%)	0 (0%)	
Ascites			**0.003**
1	56 (55%)	78 (76%)	
2	29 (28%)	18 (18%)	
3	17 (17%)	6 (5.9%)	
INR	1.37 (1.20, 1.67)	1.25 (1.16, 1.46)	**0.005**
AFP	270 (60, 2,418)	9 (5, 72)	**<0.001**
Hemoglobin	12.10 (10.80, 13.10)	13.70 (12.03, 14.90)	**<0.001**
MCV	94 (89, 100)	95 (90, 100)	0.3
Leucocytes	9 (6, 3,698)	6 (5, 13)	**0.020**
Platelets	113,500 (272, 209,796)	85,500 (202, 159,750)	0.059
Albumin	3.29 (2.98, 3.54)	3.75 (3.10, 4.20)	**<0.001**
Total_Bil	1.71 (1.02, 3.40)	1.30 (0.80, 2.38)	**0.003**
ALT	60 (33, 83)	45 (31, 71)	0.13
AST	85 (60, 124)	59 (38, 93)	**<0.001**
GGT	229 (118, 432)	156 (82, 276)	**0.006**
ALP	222 (146, 335)	130 (97, 182)	**<0.001**
TP	6.90 (6.30, 7.30)	7.15 (6.40, 7.60)	0.11
Creatinine	0.88 (0.76, 1.21)	0.80 (0.70, 1.02)	**0.022**
Nodule			0.2
0	0 (0%)	1 (1.0%)	
1	32 (31%)	46 (45%)	
2	20 (20%)	14 (14%)	
3	11 (11%)	7 (6.9%)	
4	5 (4.9%)	2 (2.0%)	
5	34 (33%)	32 (31%)	
Major_Dim	6.8 (3.7, 9.7)	4.7 (2.9, 7.3)	**0.004**
Dir_Bil	1.08 (0.50, 1.79)	0.50 (0.30, 1.00)	**<0.001**
Iron	75 (46, 106)	87 (47, 123)	0.7
Sat	30 (19, 70)	30 (15, 51)	0.2
Ferritin	452 (173, 856)	130 (60, 363)	**<0.001**

Splitting data into Test and Train:

By splitting the data set into random test and train data set will leading to potential colinearity issue.

```
split1 <- sample(c(rep(0, 0.7 * nrow(data)), rep(1, 0.3 * nrow(dt))))

train <- dt[split1 == 0, ]
test <- dt[split1== 1, ]

paste("train sample size: ", dim(train)[1])
```

```
## [1] "train sample size:  143"
```

```
paste("test sample size: ", dim(test)[1])
```

```
## [1] "test sample size:  61"
```

```
cats.fit <- glm(Class1 ~ Symptoms + Endemic + PVT + Metastasis +
                  Age + PS + Encephalopathy + Ascites + INR + AFP +
                  Hemoglobin + Leucocytes + Albumin + Total_Bil + AST + GGT +
                  ALP + Creatinine + Major_Dim + Dir_Bil + Ferritin, data=train,
              family=binomial)
summary(cats.fit)
```

```
##
## Call:
## glm(formula = Class1 ~ Symptoms + Endemic + PVT + Metastasis +
##     Age + PS + Encephalopathy + Ascites + INR + AFP + Hemoglobin +
##     Leucocytes + Albumin + Total_Bil + AST + GGT + ALP + Creatinine +
##     Major_Dim + Dir_Bil + Ferritin, family = binomial, data = train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.18480  -0.62912  -0.07091   0.51328   2.36182
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     1.314e+01  4.274e+00   3.075   0.0021 **
## Symptoms       -1.367e+00  6.002e-01  -2.278   0.0227 *
## Endemic         2.887e+00  1.362e+00   2.120   0.0340 *
## PVT            -5.469e-01  6.601e-01  -0.828   0.4074
## Metastasis     -1.300e+00  6.993e-01  -1.859   0.0631 .
## Age            -4.936e-02  2.258e-02  -2.186   0.0288 *
## PS             -3.203e-01  3.019e-01  -1.061   0.2887
## Encephalopathy -4.107e-01  6.253e-01  -0.657   0.5112
## Ascites        -2.265e-01  4.223e-01  -0.536   0.5917
## INR            -1.938e+00  8.534e-01  -2.271   0.0232 *
## AFP             2.901e-06  2.215e-06   1.310   0.1903
## Hemoglobin      4.346e-02  1.453e-01   0.299   0.7649
## Leucocytes      8.521e-06  9.517e-05   0.090   0.9287
## Albumin        -5.939e-01  5.039e-01  -1.179   0.2385
## Total_Bil       8.245e-02  2.318e-01   0.356   0.7221
```

4

```
## AST             -6.054e-03  3.795e-03  -1.595   0.1107
## GGT              1.106e-03  1.783e-03   0.621   0.5349
## ALP             -5.901e-03  2.361e-03  -2.499   0.0125 *
## Creatinine      -4.139e-01  3.265e-01  -1.268   0.2050
## Major_Dim       -9.643e-02  6.127e-02  -1.574   0.1155
## Dir_Bil         -1.286e-01  3.592e-01  -0.358   0.7203
## Ferritin        -2.046e-03  8.762e-04  -2.335   0.0195 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 198.07  on 142  degrees of freedom
## Residual deviance: 111.94  on 121  degrees of freedom
## AIC: 155.94
##
## Number of Fisher Scoring iterations: 6
```

```
check_model(cats.fit)
```
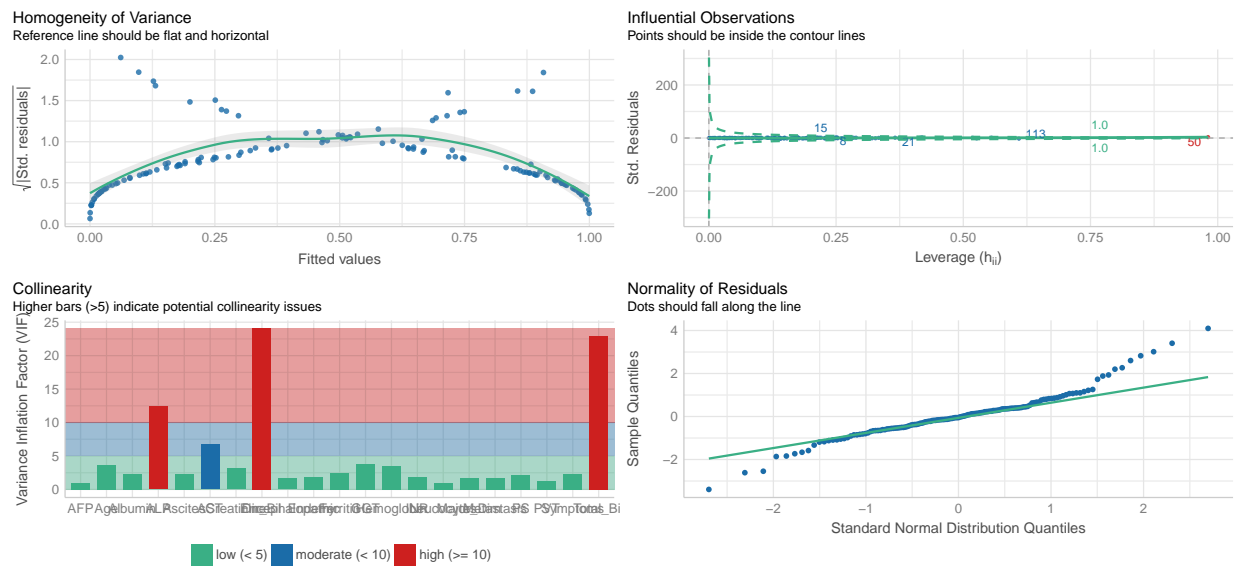

After the model is created and fitted, this model is used for making predictions on the test data set values.

```
pred <- predict(cats.fit, newdata = test, type = "response")
pred_bens <- as.factor(ifelse(pred>0.5,1,0))
```

```
confusionMatrix(data = pred_bens, reference = test$Class1)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 23 10
##          1  5 23
##
```

```
##                Accuracy : 0.7541
##                  95% CI : (0.6271, 0.8554)
##     No Information Rate : 0.541
##     P-Value [Acc > NIR] : 0.0004963
##
##                   Kappa : 0.5115
##
##  Mcnemar's Test P-Value : 0.3016996
##
##             Sensitivity : 0.8214
##             Specificity : 0.6970
##          Pos Pred Value : 0.6970
##          Neg Pred Value : 0.8214
##              Prevalence : 0.4590
##          Detection Rate : 0.3770
##    Detection Prevalence : 0.5410
##       Balanced Accuracy : 0.7592
##
##        'Positive' Class : 0
##
```
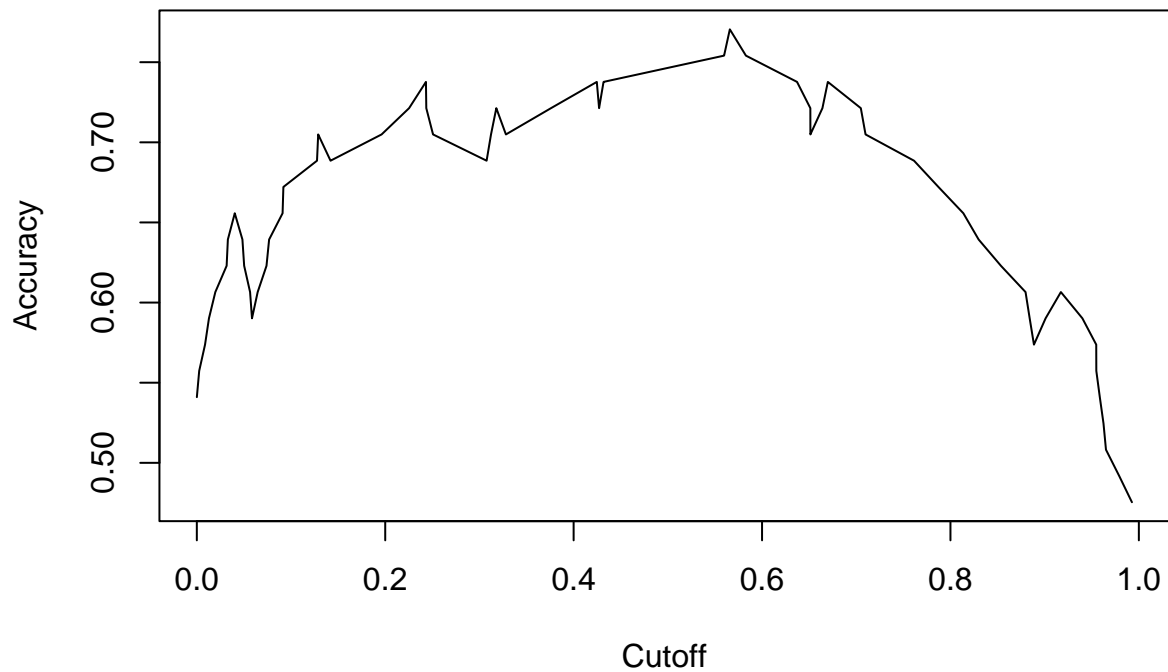
Cutoff Values vs Accuracy:

```r
pred.rocr <- ROCR::prediction(pred, test$Class1)

eval <- ROCR::performance(pred.rocr, "acc")

plot(eval)
```

Create AUC and ROC for test data:

ROC CURVE - ROC (Receiver Operator Characteristic Curve) can help in deciding the best threshold value. A ROC curve is plotted with FPR on the X-axis and TPR on the y-axis. A high threshold value gives - high specificity and low sensitivity A low threshold value gives - low specificity and high sensitivity.

```
library(pROC)
test_roc <- roc(test$Class1 ~ pred, plot = TRUE, print.auc = TRUE)
```
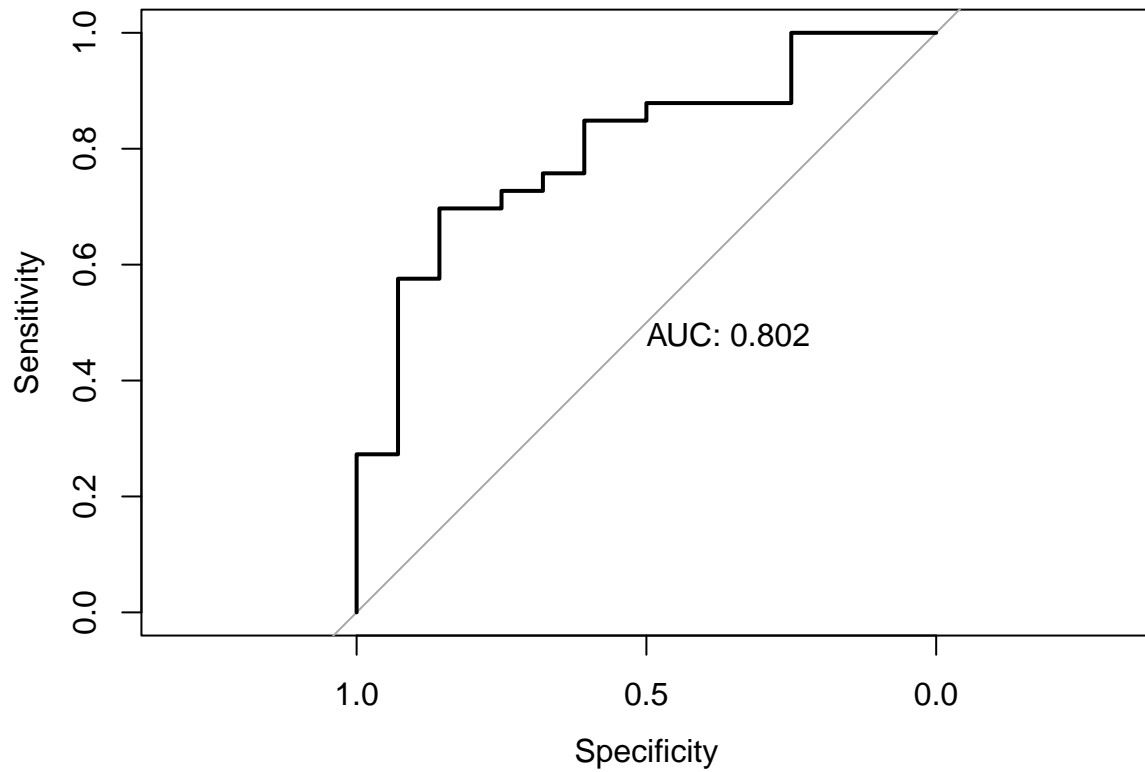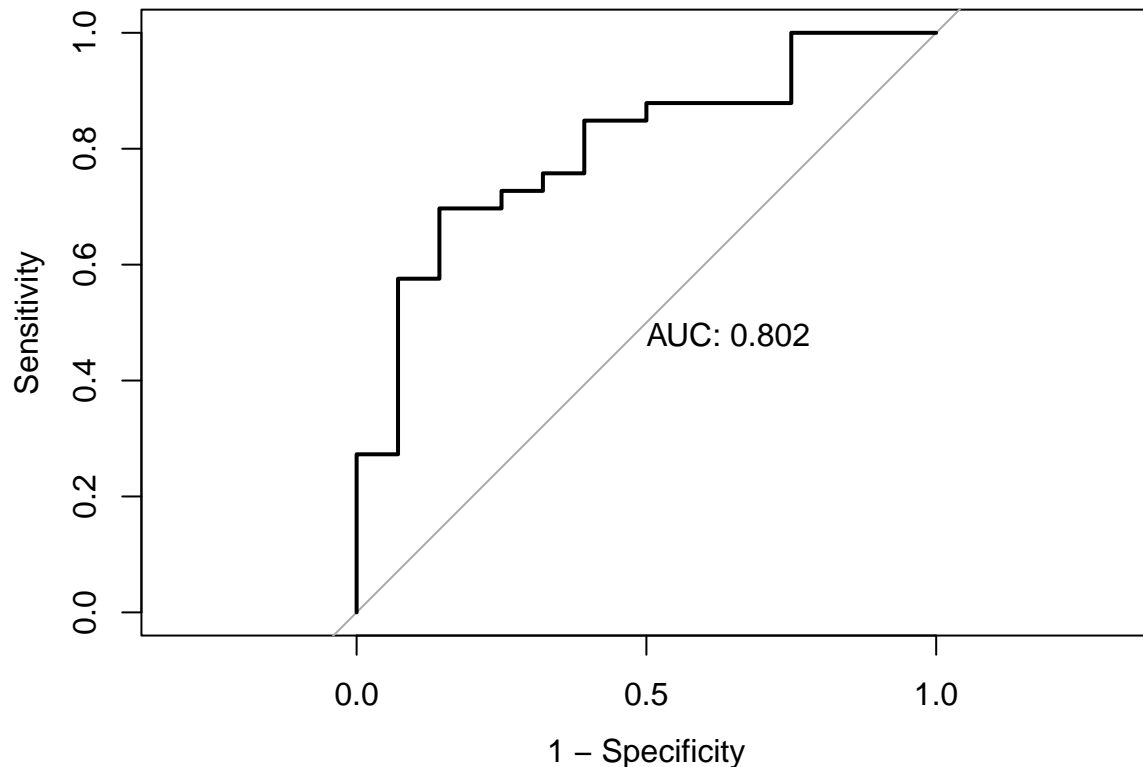
```
auc(test_roc)
```

```
## Area under the curve: 0.8019
```

```
plot(test_roc,
     print.cutoffs.at=-c(50,90,110,120),
     cutoff.label.function=`-`,
     points.pch = 20, points.col = "darkblue",
     text.adj=c(1.2,-0.5), legacy.axes = TRUE,
     print.auc = TRUE)
```

The model was verified in the external data set, AUC =0.824, indicating that the model has a good discrimination in the validation of the external data set.

As the random data splitting results in a potential colinearity issue.

Then, I will try the Bootstraping Method.

Bootstraping:

```
dt2 <- datadist(dt)

options(datadist = "dt2")

fit1 <- lrm(Class1~Symptoms + Endemic + PVT + Metastasis + Age + PS +
              Encephalopathy + Ascites + INR + AFP + Hemoglobin + Leucocytes +
              Albumin + Total_Bil + AST + GGT + ALP + Creatinine + Major_Dim +
              Dir_Bil + Ferritin, data = dt, x=T, y=T)

fit1
```

```
## Logistic Regression Model
##
##   lrm(formula = Class1 ~ Symptoms + Endemic + PVT + Metastasis +
```
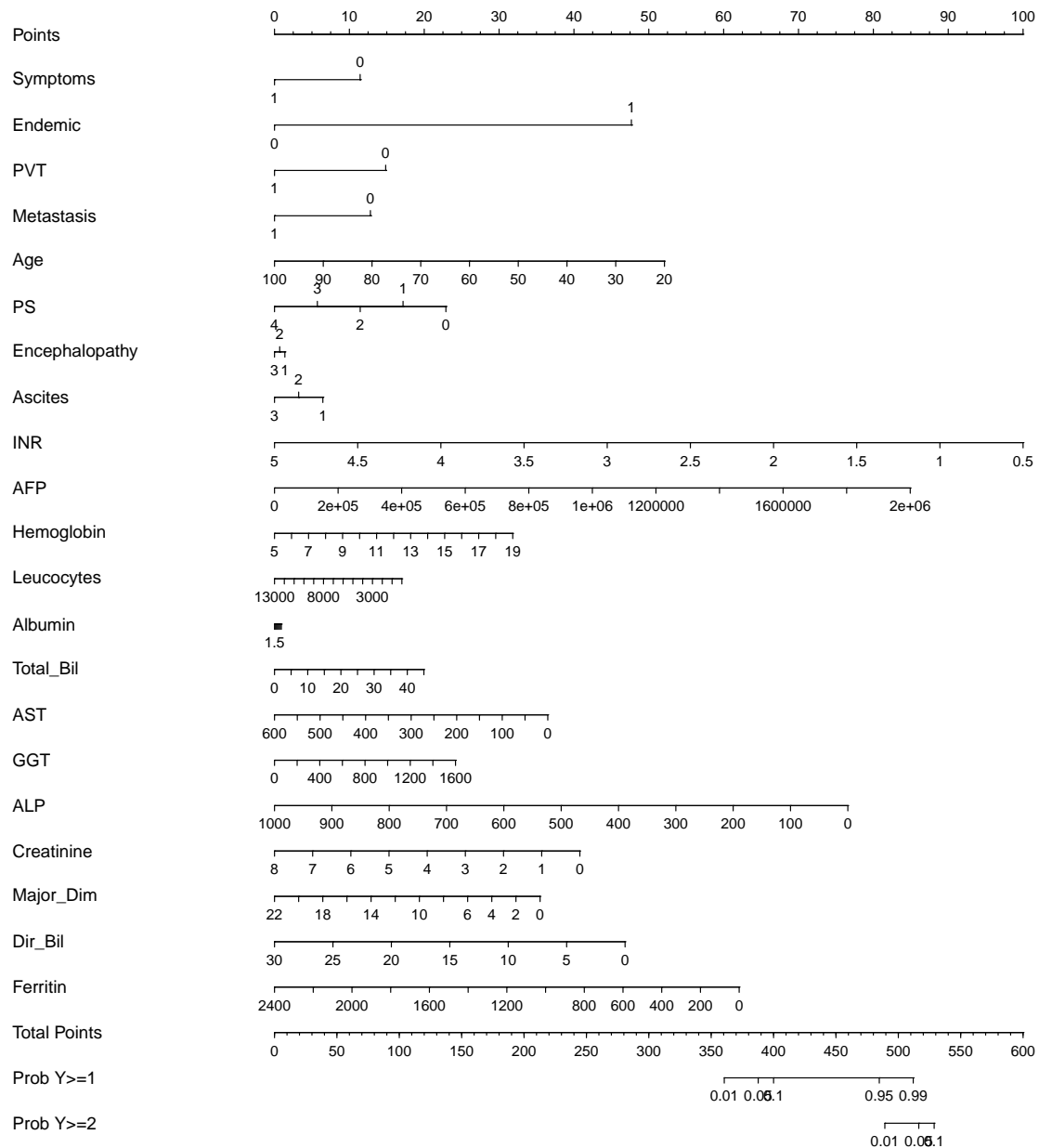
```
##        Age + PS + Encephalopathy + Ascites + INR + AFP + Hemoglobin +
##        Leucocytes + Albumin + Total_Bil + AST + GGT + ALP + Creatinine +
##        Major_Dim + Dir_Bil + Ferritin, data = dt, x = T, y = T)
##
##                        Model Likelihood       Discrimination    Rank Discrim.
##                           Ratio Test              Indexes           Indexes
##  Obs            204    LR chi2      110.67    R2       0.558    C       0.886
##   0             102    d.f.             21    g        2.712    Dxy     0.772
##   1             102    Pr(> chi2) <0.0001    gr      15.055    gamma   0.772
##  max |deriv| 2e-06                            gp       0.389    tau-a   0.388
##                                               Brier    0.140
##
##
##                 Coef     S.E.    Wald Z Pr(>|Z|)
##  Intercept      7.1159 2.8715    2.48   0.0132
##  Symptoms      -0.6939 0.4452   -1.56   0.1191
##  Endemic        2.8928 1.2514    2.31   0.0208
##  PVT           -0.9011 0.5293   -1.70   0.0887
##  Metastasis    -0.7772 0.4987   -1.56   0.1191
##  Age           -0.0395 0.0186   -2.13   0.0335
##  PS            -0.3473 0.2242   -1.55   0.1214
##  Encephalopathy -0.0420 0.5133  -0.08   0.9348
##  Ascites       -0.1959 0.3236   -0.61   0.5450
##  INR           -1.3484 0.6510   -2.07   0.0383
##  AFP            0.0000 0.0000    1.48   0.1393
##  Hemoglobin     0.1379 0.1090    1.27   0.2057
##  Leucocytes    -0.0001 0.0001   -1.01   0.3135
##  Albumin        0.0157 0.3790    0.04   0.9669
##  Total_Bil      0.0269 0.1821    0.15   0.8825
##  AST           -0.0037 0.0027   -1.37   0.1695
##  GGT            0.0009 0.0012    0.74   0.4617
##  ALP           -0.0046 0.0018   -2.59   0.0096
##  Creatinine    -0.3094 0.2400   -1.29   0.1974
##  Major_Dim     -0.0979 0.0491   -1.99   0.0463
##  Dir_Bil       -0.0947 0.2819   -0.34   0.7369
##  Ferritin      -0.0016 0.0006   -2.84   0.0045
##
```

According to the calculated results, the C-Statistics is 0.886 in this data, which indicate a strong model.

```r
fun2 <- function(x)plogis(x-fit1$coefficients[1]+fit1$coefficients[2])
fun3 <- function(x)plogis(x-fit1$coefficients[1]+fit1$coefficients[3])

nom1 <- nomogram(fit1, fun = list('Prob Y>=1' = plogis, 'Prob Y>=2' = fun2),
             fun.at = c(.01, .05, seq(.1, .9, by=1), .95, .99),
             lp = F, funlabel = "Mortality Risk from HCC", minlength = 1,
             abbrev = T)

plot(nom1)
```

Points   0   10   20   30   40   50   60   70   80   90   100

Symptoms   0 / 1   1

Endemic   0 / 1 ... 1

PVT   0 ... 0 / 1

Metastasis   0 ... 1

Age   100   90   80   70   60   50   40   30   20

PS   3   1 / 4 2   2   0

Encephalopathy   3 1 / 2

Ascites   3   1 / 2

INR   5   4.5   4   3.5   3   2.5   2   1.5   1   0.5

AFP   0   2e+05   4e+05   6e+05   8e+05   1e+06   1200000   1600000   2e+06

Hemoglobin   5   7   9   11   13   15   17   19

Leucocytes   13000   8000   3000

Albumin   1.5

Total_Bil   0   10   20   30   40

AST   600   500   400   300   200   100   0

GGT   0   400   800   1200   1600

ALP   1000   900   800   700   600   500   400   300   200   100   0

Creatinine   8   7   6   5   4   3   2   1   0

Major_Dim   22   18   14   10   6   4   2   0

Dir_Bil   30   25   20   15   10   5   0

Ferritin   2400   2000   1600   1200   800   600   400   200   0

Total Points   0   50   100   150   200   250   300   350   400   450   500   550   600

Prob Y>=1   0.01   0.05   0.1   0.95   0.99

Prob Y>=2   0.01   0.05   0.1

```r
call <- calibrate(fit1, method = "boot", B=100)
plot(call, xlim=c(0, 1.0), ylim=c(0, 1.0))
```

B= 100 repetitions, boot   Mean absolute error=0.063 n=204
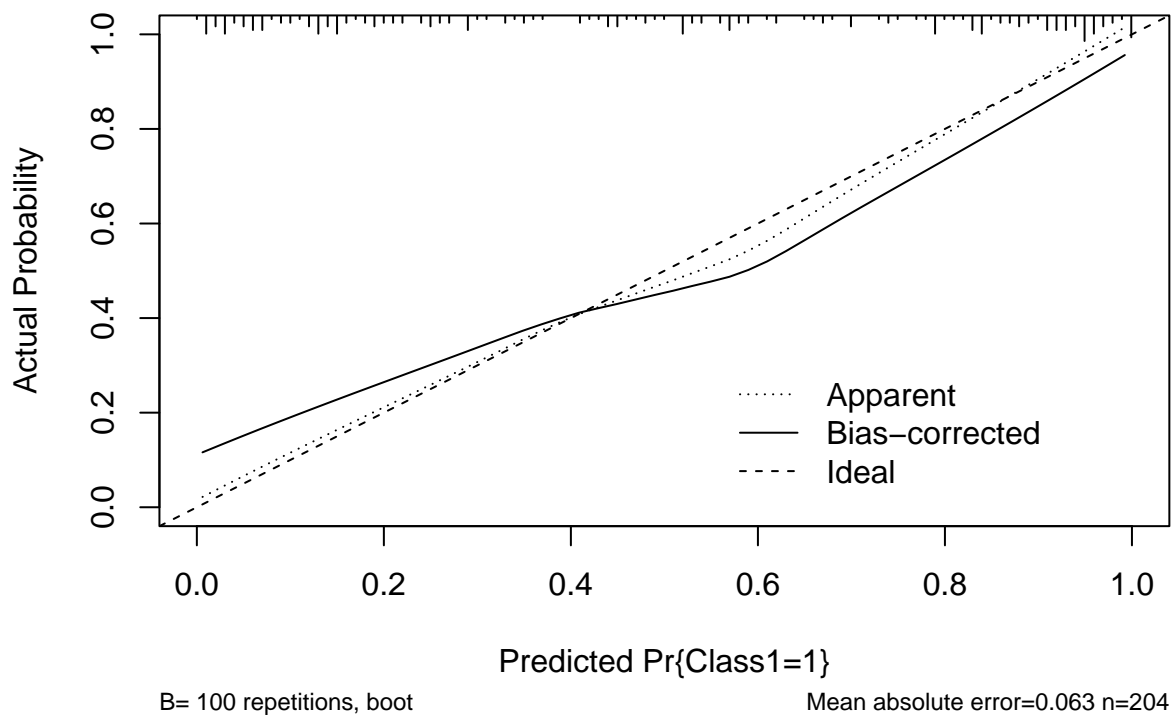
```
## 
## n=204    Mean absolute error=0.063    Mean squared error=0.00459
## 0.9 Quantile of absolute error=0.099
```

it is obvious that the contribution of some predictors to the model are negligible, like, Encephalopathy, Albumin, Total Bilirubin, and Direct Bilirubin.

Exclude these variables that contributes less to the result from the regression model, then reconstruct model "fit2" and display the model parameters.

```
fit2 <- lrm(Class1~Symptoms + Endemic + PVT + Metastasis + Age + PS + Ascites +
            INR + AFP + Hemoglobin + Leucocytes +  AST + GGT + ALP +
            Creatinine + Major_Dim + Ferritin, data = dt, x=T, y=T)

fit2
```

```
## Logistic Regression Model
## 
##  lrm(formula = Class1 ~ Symptoms + Endemic + PVT + Metastasis +
##      Age + PS + Ascites + INR + AFP + Hemoglobin + Leucocytes +
##      AST + GGT + ALP + Creatinine + Major_Dim + Ferritin, data = dt,
##      x = T, y = T)
## 
##                      Model Likelihood    Discrimination    Rank Discrim.
##                        Ratio Test           Indexes           Indexes
```

```
## Obs            204    LR chi2      110.16   R2       0.556    C      0.884
##  0             102    d.f.             17   g        2.678    Dxy    0.768
##  1             102    Pr(> chi2) <0.0001   gr      14.559    gamma  0.769
## max |deriv| 2e-08                           gp       0.388    tau-a  0.386
##                                             Brier    0.140
##
##
##             Coef   S.E.   Wald Z Pr(>|Z|)
## Intercept   7.0894 2.4005   2.95  0.0031
## Symptoms   -0.7147 0.4360  -1.64  0.1012
## Endemic     2.8795 1.2414   2.32  0.0204
## PVT        -0.8860 0.5243  -1.69  0.0911
## Metastasis -0.8165 0.4871  -1.68  0.0937
## Age        -0.0374 0.0182  -2.06  0.0396
## PS         -0.3603 0.2067  -1.74  0.0813
## Ascites    -0.2543 0.3050  -0.83  0.4044
## INR        -1.3722 0.6157  -2.23  0.0258
## AFP         0.0000 0.0000   1.53  0.1271
## Hemoglobin  0.1418 0.1050   1.35  0.1772
## Leucocytes -0.0001 0.0001  -1.09  0.2749
## AST        -0.0041 0.0026  -1.58  0.1145
## GGT         0.0009 0.0012   0.73  0.4676
## ALP        -0.0047 0.0018  -2.67  0.0076
## Creatinine -0.3284 0.2331  -1.41  0.1589
## Major_Dim  -0.0954 0.0481  -1.98  0.0472
## Ferritin   -0.0016 0.0005  -2.89  0.0038
##
```
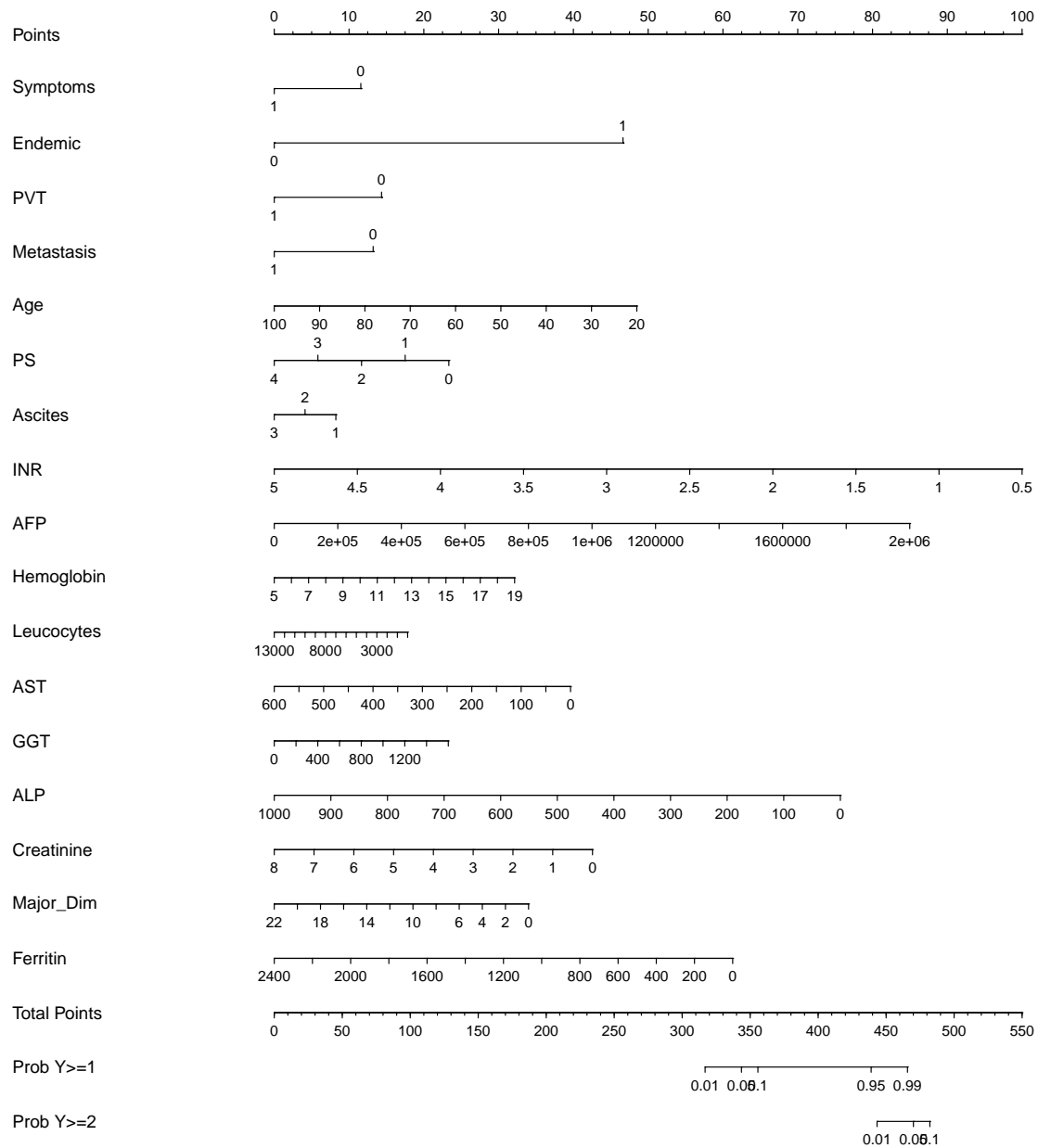
The C-Statistics = 0.884. The nomogram for the fit2 model become:

```r
fun2 <- function(x)plogis(x-fit1$coefficients[1]+fit1$coefficients[2])
fun3 <- function(x)plogis(x-fit1$coefficients[1]+fit1$coefficients[3])

nom2 <- nomogram(fit2, fun = list('Prob Y>=1' = plogis, 'Prob Y>=2' = fun2),
                 fun.at = c(.01, .05, seq(.1, .9, by=1), .95, .99),
                 lp = F, funlabel = "Mortality Risk from HCC",
                 minlength = 1, abbrev = T)

plot(nom2)
```

13

| Points | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

Symptoms
0
1

Endemic
1
0

PVT
0
1

Metastasis
0
1

Age
100 90 80 70 60 50 40 30 20

PS
3 1
4 2 0

Ascites
2
3 1

INR
5 4.5 4 3.5 3 2.5 2 1.5 1 0.5

AFP
0 2e+05 4e+05 6e+05 8e+05 1e+06 1200000 1600000 2e+06

Hemoglobin
5 7 9 11 13 15 17 19

Leucocytes
13000 8000 3000

AST
600 500 400 300 200 100 0

GGT
0 400 800 1200

ALP
1000 900 800 700 600 500 400 300 200 100 0

Creatinine
8 7 6 5 4 3 2 1 0

Major_Dim
22 18 14 10 6 4 2 0

Ferritin
2400 2000 1600 1200 800 600 400 200 0

Total Points
0 50 100 150 200 250 300 350 400 450 500 550

Prob Y>=1
0.01 0.05 0.1 0.95 0.99

Prob Y>=2
0.01 0.05 0.1

And the Calibration Model Become:

```
call2 <- calibrate(fit2, method = "boot", B=100)
plot(call2, xlim=c(0, 1.0), ylim=c(0, 1.0))
```

B= 100 repetitions, boot   Mean absolute error=0.057 n=204
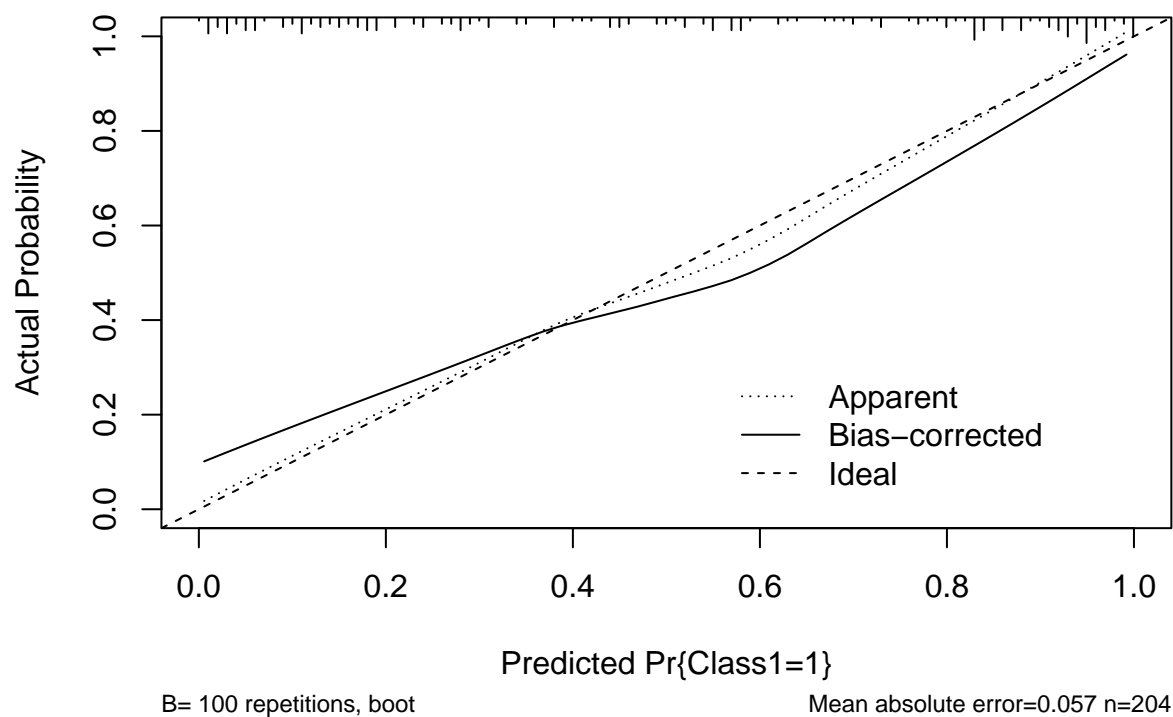
```
##
## n=204    Mean absolute error=0.057    Mean squared error=0.00374
## 0.9 Quantile of absolute error=0.086
```

Calculation C-Statistics of Logistic Regression Model:

```
dt$predvalue <- predict(fit1)

pred1 <- prediction(dt$predvalue, dt$Class1)

pref <- ROCR::performance(pred1, "tpr", "fpr")

plot(pref)
abline(0, 1, col = 3, lty = 2)
```
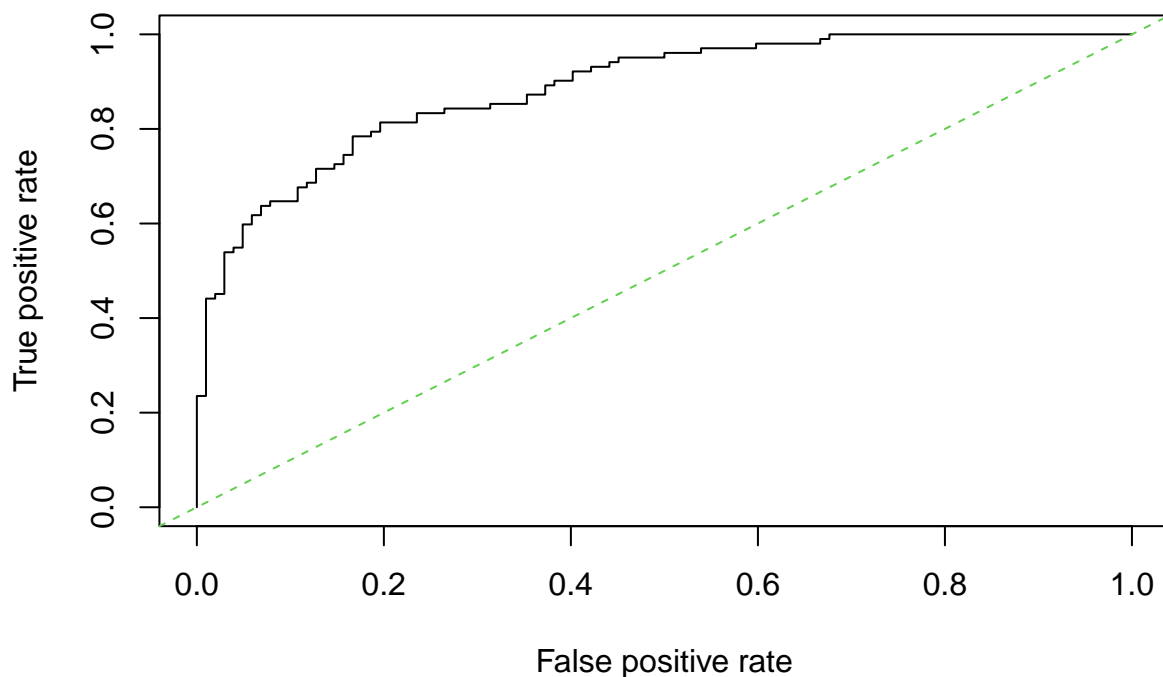
With Hmisc library:

```
library(Hmisc)

somers2(dt$predvalue, dt$Class2)
```

```
##          C        Dxy          n    Missing
##   0.8861015  0.7722030 204.0000000  0.0000000
```

Calculating the area under the ROC curve (AUC) is C-statistics = 0.8861015, which is consistent with the above calculation results.

NRI calculation of dichotomous outcome:

Build a matrix out of a subset of the data containing Symptoms, Endemic, PVT, Metastasis, Age, PS, Encephalopathy, Ascites, INR, AFP, Hemoglobin, Leucocytes, Albumin, Total Bilirubin, AST, GGT, ALP, Creatinine, Major Dimension, Direct Bilirubin, and Ferritin.

```
event <- dt$Class1

z.std <- as.matrix(subset(dt, select = c(Symptoms, Endemic, PVT, Metastasis,
                                        Age, PS, Encephalopathy, Ascites, INR,
                                        AFP, Hemoglobin, Leucocytes, Albumin,
                                        Total_Bil, AST, GGT, ALP, Creatinine,
                                        Major_Dim, Dir_Bil, Ferritin)))
```

Build a matrix out of a subset of data containing Symptoms, Endemic, PVT, Metastasis, Age, PS, Ascites, INR, AFP, Hemoglobin, Leucocytes, AST, GGT, ALP, Creatinine, Major Dimension, and Ferritin.

```
z.new <- as.matrix(subset(dt, select = c(Symptoms, Endemic, PVT, Metastasis,
                                          Age, PS, Ascites, INR, AFP, Hemoglobin,
                                          Leucocytes, AST, GGT, ALP, Creatinine,
                                          Major_Dim, Ferritin)))
```

Construct two logistic regression model: mstd and mnew. Model "mnew" has one more variable "protime". Calculation using "nricens" library requires x = TRUE, which means that output contains the matrix.

```
mstd <- glm(event ~ Symptoms + Endemic + PVT + Metastasis + Age + PS +
              Encephalopathy + Ascites + INR + AFP + Hemoglobin + Leucocytes +
              Albumin + Total_Bil + AST + GGT + ALP + Creatinine + Major_Dim +
              Dir_Bil + Ferritin, binomial(logit),
            data.frame(dt$Class1, z.std), x=T)

mnew <- glm(event ~ Symptoms + Endemic + PVT + Metastasis + Age + PS + Ascites
              + INR + AFP + Hemoglobin + Leucocytes + AST + GGT + ALP +
              Creatinine + Major_Dim + Ferritin, binomial(logit),
            data.frame(dt$Class1, z.new), x=T)
```

```
nribin(mdl.std = mstd, mdl.new = mnew, cut = c(0.2, 0.4),
       niter = 100, updown = 'category')
```

Calculating the predicted risk of two models.

```
##          New
## Standard < 0.2 < 0.4 >= 0.4
##    < 0.2     57     2      0
##    < 0.4      1    25      0
##    >= 0.4     0     3    116
##          New
## Standard < 0.2 < 0.4 >= 0.4
##    < 0.2      5     0      0
##    < 0.4      0    10      0
##    >= 0.4     0     1     86
##          New
## Standard < 0.2 < 0.4 >= 0.4
##    < 0.2     52     2      0
##    < 0.4      1    15      0
##    >= 0.4     0     2     30
```
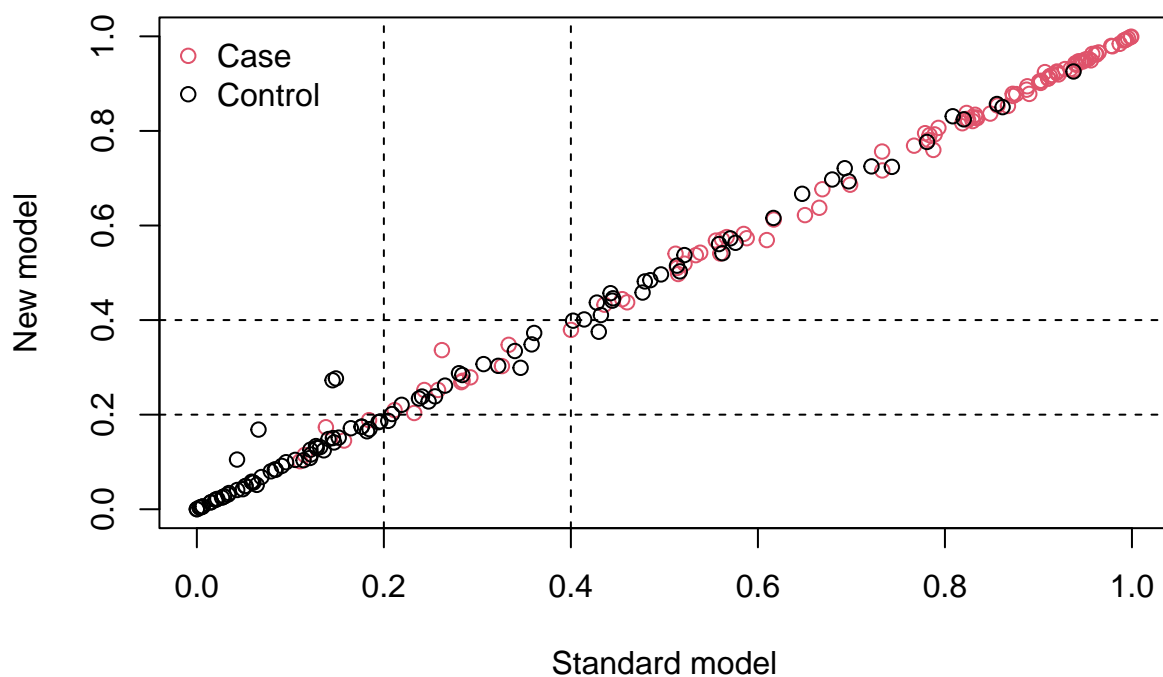
```
##                   Estimate
## NRI             0.000000000
## NRI+           -0.009803922
## NRI-            0.009803922
## Pr(Up|Case)     0.000000000
## Pr(Down|Case)   0.009803922
## Pr(Down|Ctrl)   0.029411765
## Pr(Up|Ctrl)     0.019607843
##                   Estimate   Std.Error         Lower        Upper
## NRI             0.000000000 0.05587340 -0.159167550  0.04961538
## NRI+           -0.009803922 0.02733146 -0.067961165  0.03773585
## NRI-            0.009803922 0.04470780 -0.114035088  0.03773585
## Pr(Up|Case)     0.000000000 0.01646233  0.000000000  0.05660377
## Pr(Down|Case)   0.009803922 0.02565507  0.000000000  0.09782609
## Pr(Down|Ctrl)   0.029411765 0.02932570  0.000000000  0.11000000
## Pr(Up|Ctrl)     0.019607843 0.04535467  0.009345794  0.17346939
```

```
nribin(event = dt$Class2, z.std = z.std, z.new = z.new, cut = c(0.2, 0.4),
       niter = 100, updown = 'category')
```

Calculation of risk difference NRI using ('event', 'z.std', 'z.new')

```
##                      Estimate    Std. Error     z value    Pr(>|z|)
```

```
## (Intercept)      7.115906e+00 2.871519e+00  2.47809795 0.013208488
## Symptoms       -6.939419e-01 4.451858e-01 -1.55876936 0.119050978
## Endemic         2.892767e+00 1.251439e+00  2.31155280 0.020802340
## PVT            -9.010988e-01 5.292769e-01 -1.70250909 0.088659976
## Metastasis     -7.771897e-01 4.986863e-01 -1.55847432 0.119120852
## Age            -3.951218e-02 1.858395e-02 -2.12614570 0.033491130
## PS             -3.472762e-01 2.242129e-01 -1.54886812 0.121413426
## Encephalopathy -4.196384e-02 5.133446e-01 -0.08174595 0.934848741
## Ascites        -1.958771e-01 3.235797e-01 -0.60534441 0.544950179
## INR            -1.348423e+00 6.509763e-01 -2.07138508 0.038322824
## AFP             2.576500e-06 1.742743e-06  1.47841630 0.139296386
## Hemoglobin      1.379279e-01 1.089844e-01  1.26557523 0.205665199
## Leucocytes     -7.940936e-05 7.879173e-05 -1.00783872 0.313531893
## Albumin         1.574888e-02 3.790132e-01  0.04155234 0.966855569
## Total_Bil       2.690795e-02 1.820868e-01  0.14777541 0.882520015
## AST            -3.691803e-03 2.687179e-03 -1.37385851 0.169485611
## GGT             9.169550e-04 1.245702e-03  0.73609483 0.461672991
## ALP            -4.647349e-03 1.793620e-03 -2.59104363 0.009568536
## Creatinine     -3.093852e-01 2.400035e-01 -1.28908630 0.197368086
## Major_Dim      -9.785202e-02 4.910300e-02 -1.99279117 0.046284319
## Dir_Bil        -9.471961e-02 2.819262e-01 -0.33597305 0.736891186
## Ferritin       -1.569376e-03 5.526192e-04 -2.83988686 0.004512954
##                   Estimate   Std. Error    z value     Pr(>|z|)
## (Intercept)  7.089440e+00 2.400407e+00  2.9534318 0.003142620
## Symptoms    -7.147155e-01 4.359751e-01 -1.6393495 0.101140489
## Endemic      2.879497e+00 1.241203e+00  2.3199253 0.020344918
## PVT         -8.860239e-01 5.243127e-01 -1.6898769 0.091051512
## Metastasis  -8.165420e-01 4.870997e-01 -1.6763345 0.093672698
## Age         -3.742097e-02 1.818751e-02 -2.0575088 0.039637306
## PS          -3.603437e-01 2.067357e-01 -1.7430167 0.081330699
## Ascites     -2.542969e-01 3.049926e-01 -0.8337807 0.404404600
## INR         -1.372158e+00 6.156501e-01 -2.2287951 0.025827543
## AFP          2.623552e-06 1.719439e-06  1.5258186 0.127055049
## Hemoglobin   1.417581e-01 1.050356e-01  1.3496199 0.177137938
## Leucocytes  -8.471347e-05 7.758664e-05 -1.0918564 0.274896229
## AST         -4.077217e-03 2.583308e-03 -1.5782930 0.114498313
## GGT          8.980442e-04 1.236181e-03  0.7264665 0.467552867
## ALP         -4.674011e-03 1.750622e-03 -2.6699149 0.007587046
## Creatinine  -3.284470e-01 2.331042e-01 -1.4090137 0.158831128
## Major_Dim   -9.544225e-02 4.808990e-02 -1.9846632 0.047181955
## Ferritin    -1.577167e-03 5.448224e-04 -2.8948277 0.003793667
##         New
## Standard < 0.2 < 0.4 >= 0.4
##   < 0.2     57     2      0
##   < 0.4      1    25      0
##   >= 0.4     0     3    116
##         New
## Standard < 0.2 < 0.4 >= 0.4
##   < 0.2      5     0      0
##   < 0.4      0    10      0
##   >= 0.4     0     1     86
##         New
## Standard < 0.2 < 0.4 >= 0.4
##   < 0.2     52     2      0
```

```
##   < 0.4      1     15      0
##   >= 0.4     0      2     30
```
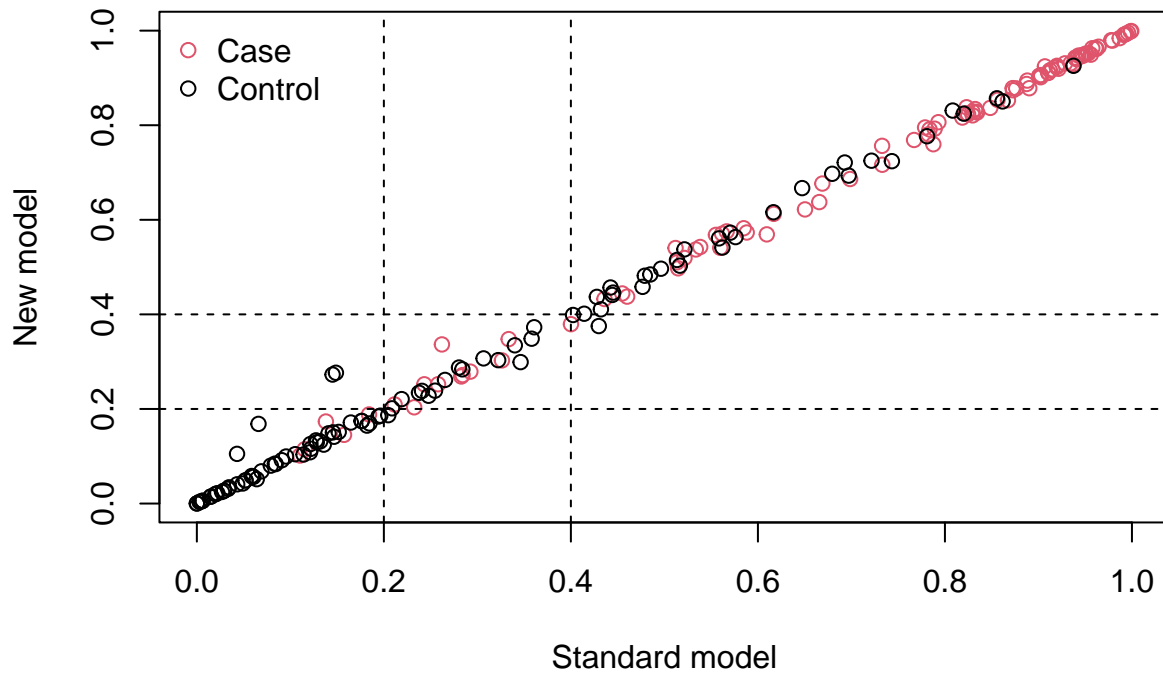


```
##                Estimate
## NRI          0.000000000
## NRI+        -0.009803922
## NRI-         0.009803922
## Pr(Up|Case)  0.000000000
## Pr(Down|Case) 0.009803922
## Pr(Down|Ctrl) 0.029411765
## Pr(Up|Ctrl)  0.019607843
##                Estimate  Std.Error        Lower        Upper
## NRI          0.000000000 0.05448640 -0.155000000 0.06281915
## NRI+        -0.009803922 0.02631921 -0.070175439 0.02830189
## NRI-         0.009803922 0.04208473 -0.116504854 0.06451613
## Pr(Up|Case)  0.000000000 0.01854950  0.000000000 0.06593407
## Pr(Down|Case) 0.009803922 0.02215811  0.000000000 0.08080808
## Pr(Down|Ctrl) 0.029411765 0.02679062  0.009174312 0.10416667
## Pr(Up|Ctrl)  0.019607843 0.04217346  0.009523810 0.17708333
```

```
nribin(mdl.std = mstd, mdl.new = mnew, cut = 0.02, niter = 0, updown = 'diff')
```
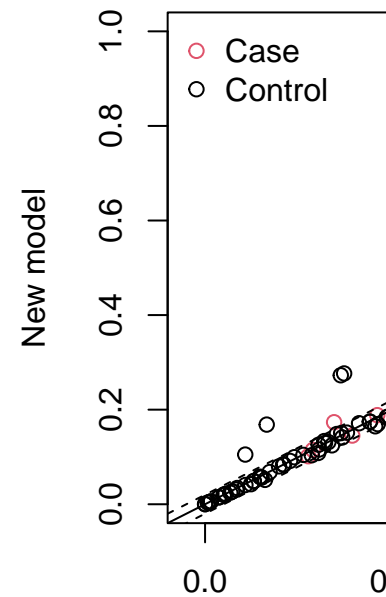
Calculation of risk difference NRI using ('mdl.std', 'mdl.new'), updown = 'diff':

```
##                      Estimate
## NRI               -0.05882353
## NRI+              -0.03921569
## NRI-              -0.01960784
## Pr(Up|Case)        0.03921569
## Pr(Down|Case)      0.07843137
## Pr(Down|Ctrl)      0.03921569
## Pr(Up|Ctrl)        0.05882353
```

```
nribin(event = dt$Class2, z.std = z.std, z.new = z.new, cut = 0.02,
       niter = 100, updown = 'diff')
```

Calculation of risk difference NRI using ('event', 'z.std', 'z.new'), updown = 'diff':

```
##                     Estimate    Std. Error      z value     Pr(>|z|)
## (Intercept)     7.115906e+00  2.871519e+00   2.47809795  0.013208488
## Symptoms       -6.939419e-01  4.451858e-01  -1.55876936  0.119050978
## Endemic         2.892767e+00  1.251439e+00   2.31155280  0.020802340
## PVT            -9.010988e-01  5.292769e-01  -1.70250909  0.088659976
## Metastasis     -7.771897e-01  4.986863e-01  -1.55847432  0.119120852
## Age            -3.951218e-02  1.858395e-02  -2.12614570  0.033491130
## PS             -3.472762e-01  2.242129e-01  -1.54886812  0.121413426
```

```
## Encephalopathy -4.196384e-02 5.133446e-01 -0.08174595 0.934848741
## Ascites        -1.958771e-01 3.235797e-01 -0.60534441 0.544950179
## INR            -1.348423e+00 6.509763e-01 -2.07138508 0.038322824
## AFP             2.576500e-06 1.742743e-06  1.47841630 0.139296386
## Hemoglobin      1.379279e-01 1.089844e-01  1.26557523 0.205665199
## Leucocytes     -7.940936e-05 7.879173e-05 -1.00783872 0.313531893
## Albumin         1.574888e-02 3.790132e-01  0.04155234 0.966855569
## Total_Bil       2.690795e-02 1.820868e-01  0.14777541 0.882520015
## AST            -3.691803e-03 2.687179e-03 -1.37385851 0.169485611
## GGT             9.169550e-04 1.245702e-03  0.73609483 0.461672991
## ALP            -4.647349e-03 1.793620e-03 -2.59104363 0.009568536
## Creatinine     -3.093852e-01 2.400035e-01 -1.28908630 0.197368086
## Major_Dim      -9.785202e-02 4.910300e-02 -1.99279117 0.046284319
## Dir_Bil        -9.471961e-02 2.819262e-01 -0.33597305 0.736891186
## Ferritin       -1.569376e-03 5.526192e-04 -2.83988686 0.004512954
##                    Estimate   Std. Error    z value    Pr(>|z|)
## (Intercept)  7.089440e+00 2.400407e+00  2.9534318 0.003142620
## Symptoms    -7.147155e-01 4.359751e-01 -1.6393495 0.101140489
## Endemic      2.879497e+00 1.241203e+00  2.3199253 0.020344918
## PVT         -8.860239e-01 5.243127e-01 -1.6898769 0.091051512
## Metastasis  -8.165420e-01 4.870997e-01 -1.6763345 0.093672698
## Age         -3.742097e-02 1.818751e-02 -2.0575088 0.039637306
## PS          -3.603437e-01 2.067357e-01 -1.7430167 0.081330699
## Ascites     -2.542969e-01 3.049926e-01 -0.8337807 0.404404600
## INR         -1.372158e+00 6.156501e-01 -2.2287951 0.025827543
## AFP          2.623552e-06 1.719439e-06  1.5258186 0.127055049
## Hemoglobin   1.417581e-01 1.050356e-01  1.3496199 0.177137938
## Leucocytes  -8.471347e-05 7.758664e-05 -1.0918564 0.274896229
## AST         -4.077217e-03 2.583308e-03 -1.5782930 0.114498313
## GGT          8.980442e-04 1.236181e-03  0.7264665 0.467552867
## ALP         -4.674011e-03 1.750622e-03 -2.6699149 0.007587046
## Creatinine  -3.284470e-01 2.331042e-01 -1.4090137 0.158831128
## Major_Dim   -9.544225e-02 4.808990e-02 -1.9846632 0.047181955
## Ferritin    -1.577167e-03 5.448224e-04 -2.8948277 0.003793667
```

```
##                   Estimate
## NRI            -0.05882353
## NRI+           -0.03921569
## NRI-           -0.01960784
## Pr(Up|Case)     0.03921569
## Pr(Down|Case)   0.07843137
## Pr(Down|Ctrl)   0.03921569
## Pr(Up|Ctrl)     0.05882353
##                   Estimate  Std.Error        Lower       Upper
## NRI            -0.05882353 0.12873058 -0.47976545  0.01157407
## NRI+           -0.03921569 0.08683360 -0.27184466  0.07000000
## NRI-           -0.01960784 0.08945039 -0.27777778  0.04210526
## Pr(Up|Case)     0.03921569 0.05967443  0.05434783  0.28089888
## Pr(Down|Case)   0.07843137 0.10656068  0.07207207  0.48543689
## Pr(Down|Ctrl)   0.03921569 0.06749205  0.05263158  0.32291667
## Pr(Up|Ctrl)     0.05882353 0.10307823  0.08181818  0.47422680
```

IDI calculation of dichotomous outcome:

```
library(PredictABEL)

p.std <- mstd$fitted.values
p.new <- mnew$fitted.values
```

```
dat_new <- cbind(dt, event)

reclassification(data = dat_new, cOutcome = 20, predrisk1 = p.std,
                 predrisk2 = p.new,
                 cutoff = c(0, 0.2, 0.4, 1))
```

```
##   ----------------------------------------
##
##      Reclassification table
##   ----------------------------------------
##
##   Outcome: absent
##
##              Updated Model
## Initial Model [0,0.2) [0.2,0.4) [0.4,1]  % reclassified
##     [0,0.2)        13         0       0                0
##     [0.2,0.4)       0        10       0                0
##     [0.4,1]         0         0      42                0
##
##
##   Outcome: present
##
##              Updated Model
## Initial Model [0,0.2) [0.2,0.4) [0.4,1]  % reclassified
##     [0,0.2)        44         2       0                4
##     [0.2,0.4)       1        15       0                6
##     [0.4,1]         0         3      74                4
##
##
##   Combined Data
##
##              Updated Model
## Initial Model [0,0.2) [0.2,0.4) [0.4,1]  % reclassified
##     [0,0.2)        57         2       0                3
##     [0.2,0.4)       1        25       0                4
##     [0.4,1]         0         3     116                3
##   ----------------------------------------
##
##   NRI(Categorical) [95% CI]: -0.0144 [ -0.0488 - 0.0201 ] ; p-value: 0.41309
##   NRI(Continuous) [95% CI]: 0.0201 [ -0.2688 - 0.3091 ] ; p-value: 0.89131
##   IDI [95% CI]: 0.0039 [ -6e-04 - 0.0084 ] ; p-value: 0.09244
```

IDI is 0.0039, indicating that new model improves 0.39% comparing to original model.

Decision Curve Analysis for Binary Outcome:

We can view the data points on the complex model curve, where NB can also be changed to sNB, indicating a standardized prevalence.

```
library(rmda)

complex <- decision_curve(Class2~Symptoms + Endemic + PVT + Metastasis +
```

```
                          Age + PS + Encephalopathy + Ascites + INR + AFP +
                          Hemoglobin + Leucocytes + Albumin + Total_Bil + AST
                        + GGT + ALP + Creatinine + Major_Dim + Dir_Bil +
                          Ferritin, data = dt,
                        family = binomial(link = 'logit'),
                        thresholds = seq(0, 1, by = 0.01),
                        confidence.intervals = 0.95,
                        study.design = "cohort",
                        population.prevalence = 0.3)

summary(complex, measure = 'NB')
```

```
##
## Net Benefit (95% Confidence Intervals):
## ------------------------------------------------------------------------------
```

risk threshold	cost:benefit ratio	percent high risk	All	Class2 ~ Symptoms + Endemic + PVT + Metastasis + Age + PS + Encephalopathy + Ascites + INR + AFP + Hemoglobin + Leucocytes + Albumin + Total_Bil + AST + GGT + ALP + Creatinine + Major_Dim + Dir_Bil + Ferritin
0	0:1	100 (100, 100)	0.3 (0.3, 0.3)	0.3 (0.3, 0.3)
0.01	1:99	92.451 (69.826, 94.167)	0.293 (0.293, 0.293)	0.294 (0.289, 0.296)
0.02	1:49	87.647 (63.373, 88.989)	0.286 (0.286, 0.286)	0.288 (0.282, 0.293)
0.03	3:97	82.157 (61.468, 84.6)	0.278 (0.278, 0.278)	0.284 (0.276, 0.29)
0.04	1:24	79.412 (58.125, 81.223)	0.271 (0.271, 0.271)	0.279 (0.27, 0.286)
0.05	1:19	77.353 (55.352, 77.79)	0.263 (0.263, 0.263)	0.275 (0.265, 0.283)
0.06	3:47	72.647 (53.229, 74.869)	0.255 (0.255, 0.255)	0.267 (0.26, 0.281)
0.07	7:93	67.549 (50.782, 72.251)	0.247 (0.247, 0.247)	0.262 (0.254, 0.279)
0.08	2:23	65.882 (48.868, 69.296)	0.239 (0.239, 0.239)	0.256 (0.249, 0.277)
0.09	9:91	62.843 (47.361, 67.535)	0.231 (0.231, 0.231)	0.251 (0.244, 0.275)

```
##
##     0.1         1:9             60.784              0.222               0.249
##                             (46.076, 66.404)    (0.222, 0.222)      (0.239, 0.274)
##
##    0.11        11:89            59.118              0.213               0.244
##                             (44.792, 64.412)    (0.213, 0.213)      (0.235, 0.272)
##
##    0.12         3:22            57.451              0.205               0.239
##                             (43.135, 63.158)    (0.205, 0.205)      (0.228, 0.269)
##
##    0.13        13:87            55.49               0.195               0.231
##                             (42.029, 61.394)    (0.195, 0.195)      (0.225, 0.267)
##
##    0.14         7:43            54.51               0.186               0.226
##                             (41.174, 59.953)    (0.186, 0.186)      (0.22, 0.266)
##
##    0.15         3:17            52.549              0.176               0.219
##                             (40.245, 57.984)    (0.176, 0.176)      (0.214, 0.265)
##
##    0.16         4:21            51.569              0.167               0.213
##                             (39.474, 56.193)    (0.167, 0.167)      (0.211, 0.264)
##
##    0.17        17:83            50.882              0.157               0.211
##                             (38.715, 55.12)     (0.157, 0.157)      (0.207, 0.264)
##
##    0.18         9:41            50.294              0.146               0.202
##                             (38.182, 53.133)    (0.146, 0.146)      (0.203, 0.261)
##
##    0.19        19:81            48.922              0.136               0.201
##                             (37.443, 51.857)    (0.136, 0.136)      (0.201, 0.259)
##
##     0.2         1:4             47.549              0.125               0.201
##                             (36.984, 50.35)     (0.125, 0.125)      (0.196, 0.257)
##
##    0.21        21:79            47.549              0.114               0.198
##                             (36.127, 49.161)    (0.114, 0.114)      (0.195, 0.256)
##
##    0.22        11:39            47.549              0.103               0.194
##                             (35.878, 48.641)    (0.103, 0.103)      (0.189, 0.255)
##
##    0.23        23:77            46.569              0.091               0.189
##                             (34.63, 47.257)     (0.091, 0.091)      (0.189, 0.254)
##
##    0.24         6:19            45.882              0.079               0.188
##                             (33.981, 46.54)     (0.079, 0.079)      (0.183, 0.255)
##
##    0.25         1:3             43.529              0.067               0.188
##                             (33.811, 46.119)    (0.067, 0.067)       (0.18, 0.25)
##
##    0.26        13:37            41.471              0.054               0.192
##                             (33.58, 44.792)     (0.054, 0.054)      (0.177, 0.249)
##
##    0.27        27:73            40.882              0.041               0.183
##                             (32.615, 43.655)    (0.041, 0.041)      (0.172, 0.248)
```

```
##
##    0.28         7:18           40.882              0.028                 0.18
##                           (32.364, 43.077)    (0.028, 0.028)        (0.169, 0.247)
##
##    0.29        29:71           38.824              0.014                0.185
##                           (32.182, 42.016)    (0.014, 0.014)        (0.164, 0.243)
##
##     0.3         3:7           38.137               0                   0.185
##                           (31.455, 41.322)        (0, 0)            (0.162, 0.244)
##
##    0.31        31:69          38.137              -0.014               0.182
##                           (30.882, 40.184)   (-0.014, -0.014)       (0.158, 0.24)
##
##    0.32         8:17          34.902              -0.029               0.177
##                           (30.446, 39.333)   (-0.029, -0.029)       (0.152, 0.241)
##
##    0.33        33:67          34.608              -0.045               0.172
##                           (29.898, 38.824)   (-0.045, -0.045)       (0.151, 0.24)
##
##    0.34        17:33          34.314              -0.061               0.166
##                           (29.342, 37.985)   (-0.061, -0.061)       (0.149, 0.239)
##
##    0.35         7:13           34.02              -0.077               0.161
##                           (29.019, 37.255)   (-0.077, -0.077)       (0.147, 0.238)
##
##    0.36         9:16          31.765              -0.094               0.157
##                           (28.565, 36.38)    (-0.094, -0.094)       (0.143, 0.237)
##
##    0.37        37:63          30.392              -0.111               0.162
##                           (28.039, 35.599)   (-0.111, -0.111)       (0.141, 0.236)
##
##    0.38        19:31          29.804              -0.129               0.154
##                           (27.765, 35.196)   (-0.129, -0.129)       (0.138, 0.235)
##
##    0.39        39:61          29.804              -0.148               0.152
##                           (27.152, 34.771)   (-0.148, -0.148)       (0.134, 0.234)
##
##     0.4         2:3           29.804              -0.167               0.149
##                           (26.35, 34.051)    (-0.167, -0.167)       (0.133, 0.23)
##
##    0.41        41:59          28.529              -0.186               0.146
##                           (25.807, 33.624)   (-0.186, -0.186)       (0.132, 0.232)
##
##    0.42        21:29          28.529              -0.207               0.143
##                           (25.464, 32.914)   (-0.207, -0.207)       (0.132, 0.228)
##
##    0.43        43:57          28.529              -0.228               0.141
##                           (25.115, 32.48)    (-0.228, -0.228)       (0.131, 0.229)
##
##    0.44        11:14          28.529               -0.25               0.138
##                           (24.87, 32.174)     (-0.25, -0.25)        (0.13, 0.227)
##
##    0.45         9:11          27.549              -0.273               0.138
##                           (24.65, 31.837)    (-0.273, -0.273)       (0.125, 0.224)
```

```
##
##    0.46        23:27          27.549             -0.296                0.136
##                           (24.246, 31.496)    (-0.296, -0.296)      (0.123, 0.225)
##
##    0.47        47:53          26.961             -0.321                0.127
##                           (23.763, 31.067)    (-0.321, -0.321)      (0.118, 0.225)
##
##    0.48        12:13          26.275             -0.346                0.131
##                           (23.605, 30.462)    (-0.346, -0.346)      (0.117, 0.225)
##
##    0.49        49:51          26.275             -0.373                0.128
##                           (22.954, 30.157)    (-0.373, -0.373)      (0.115, 0.225)
##
##    0.5          1:1           24.608             -0.4                  0.136
##                           (22.639, 29.628)     (-0.4, -0.4)         (0.112, 0.225)
##
##    0.51        51:49          24.608             -0.429                0.134
##                           (21.832, 29.27)     (-0.429, -0.429)       (0.11, 0.222)
##
##    0.52        13:12          24.608             -0.458                0.132
##                           (21.814, 28.846)    (-0.458, -0.458)      (0.107, 0.221)
##
##    0.53        53:47          23.922             -0.489                0.137
##                           (21.511, 28.5)      (-0.489, -0.489)      (0.102, 0.22)
##
##    0.54        27:23          23.922             -0.522                0.135
##                           (21.1, 28.263)      (-0.522, -0.522)      (0.101, 0.219)
##
##    0.55        11:9           23.333             -0.556                0.127
##                           (20.408, 27.888)    (-0.556, -0.556)      (0.098, 0.217)
##
##    0.56        14:11          22.647             -0.591                0.133
##                           (20.121, 27.389)    (-0.591, -0.591)      (0.096, 0.214)
##
##    0.57        57:43          22.647             -0.628                0.131
##                           (19.611, 27.104)    (-0.628, -0.628)       (0.09, 0.21)
##
##    0.58        29:21          22.647             -0.667                0.128
##                           (19.272, 26.948)    (-0.667, -0.667)      (0.088, 0.206)
##
##    0.59        59:41          22.353             -0.707                0.123
##                           (18.753, 26.812)    (-0.707, -0.707)      (0.086, 0.203)
##
##    0.6          3:2           22.353             -0.75                 0.121
##                           (18.329, 26.304)     (-0.75, -0.75)       (0.084, 0.204)
##
##    0.61        61:39          20.784             -0.795                0.12
##                           (17.861, 25.926)    (-0.795, -0.795)      (0.078, 0.205)
##
##    0.62        31:19          20.196             -0.842                0.112
##                           (17.728, 25.758)    (-0.842, -0.842)      (0.075, 0.205)
##
##    0.63        63:37          19.902             -0.892                0.106
##                           (17.07, 25.553)     (-0.892, -0.892)      (0.078, 0.203)
```

```
##
##    0.64         16:9          19.902              -0.944                 0.104
##                          (16.769, 25.05)     (-0.944, -0.944)       (0.074, 0.201)
##
##    0.65         13:7          19.216                -1                   0.114
##                           (16.28, 24.988)       (-1, -1)            (0.071, 0.2)
##
##    0.66        33:17          18.922              -1.059                 0.108
##                           (16.04, 24.663)    (-1.059, -1.059)       (0.07, 0.199)
##
##    0.67        67:33          17.353              -1.121                 0.111
##                          (15.492, 24.389)    (-1.121, -1.121)       (0.066, 0.201)
##
##    0.68         17:8          16.765              -1.188                 0.103
##                          (14.897, 24.076)    (-1.188, -1.188)       (0.064, 0.198)
##
##    0.69        69:31          15.882              -1.258                 0.092
##                          (14.842, 24.021)    (-1.258, -1.258)       (0.062, 0.199)
##
##     0.7          7:3          15.882              -1.333                 0.09
##                          (14.487, 23.53)     (-1.333, -1.333)       (0.059, 0.194)
##
##    0.71        71:29          15.588              -1.414                 0.085
##                           (14.037, 23.4)     (-1.414, -1.414)       (0.057, 0.191)
##
##    0.72         18:7          14.608               -1.5                  0.097
##                           (13.26, 23.035)      (-1.5, -1.5)         (0.053, 0.19)
##
##    0.73        73:27          13.922              -1.593                 0.114
##                          (12.926, 22.846)    (-1.593, -1.593)       (0.048, 0.189)
##
##    0.74        37:13          13.627              -1.692                 0.11
##                          (12.124, 22.707)    (-1.692, -1.692)       (0.047, 0.188)
##
##    0.75          3:1          13.039               -1.8                  0.103
##                          (11.662, 22.343)      (-1.8, -1.8)         (0.039, 0.188)
##
##    0.76         19:6          12.745              -1.917                 0.099
##                          (11.134, 22.015)    (-1.917, -1.917)       (0.041, 0.185)
##
##    0.77        77:23          12.745              -2.043                 0.098
##                          (10.693, 21.672)    (-2.043, -2.043)       (0.039, 0.183)
##
##    0.78        39:11          11.863              -2.182                 0.087
##                          (10.198, 21.447)    (-2.182, -2.182)       (0.034, 0.182)
##
##    0.79        79:21          11.863              -2.333                 0.086
##                           (9.722, 21.25)     (-2.333, -2.333)       (0.026, 0.181)
##
##     0.8          4:1          10.98                -2.5                  0.075
##                           (9.569, 21.181)      (-2.5, -2.5)         (0.019, 0.179)
##
##    0.81        81:19          10.686              -2.684                 0.071
##                            (8.66, 20.602)    (-2.684, -2.684)       (0.014, 0.176)
```

```
##
##    0.82        41:9           9.804                -2.889                 0.06
##                          (8.333, 20.315)       (-2.889, -2.889)       (0.011, 0.173)
##
##    0.83        83:17          9.51                 -3.118                0.055
##                          (7.826, 19.946)       (-3.118, -3.118)       (0.004, 0.175)
##
##    0.84        21:4           8.627                -3.375                0.043
##                          (7.263, 19.51)        (-3.375, -3.375)       (-0.007, 0.169)
##
##    0.85        17:3           8.333                -3.667                0.038
##                          (6.323, 19.485)       (-3.667, -3.667)       (-0.016, 0.164)
##
##    0.86        43:7           8.039                 -4                   0.031
##                           (6, 19.162)            (-4, -4)             (-0.019, 0.164)
##
##    0.87        87:13          6.471                -4.385                0.065
##                          (5.51, 18.464)        (-4.385, -4.385)       (-0.025, 0.158)
##
##    0.88        22:3           5.882                -4.833                0.059
##                          (5.243, 17.908)       (-4.833, -4.833)       (-0.039, 0.156)
##
##    0.89        89:11           5                   -5.364                0.05
##                          (4.615, 17.129)       (-5.364, -5.364)       (-0.051, 0.151)
##
##     0.9         9:1           4.118                 -6                   0.041
##                          (4.322, 16.514)        (-6, -6)             (-0.06, 0.148)
##
##    0.91        91:9           3.529                -6.778                0.035
##                          (3.169, 15.978)       (-6.778, -6.778)       (-0.078, 0.143)
##
##    0.92        23:2           2.941                -7.75                 0.029
##                          (2.872, 15.288)       (-7.75, -7.75)         (-0.079, 0.138)
##
##    0.93        93:7           2.647                 -9                   0.026
##                          (2.581, 14.842)        (-9, -9)             (-0.094, 0.133)
##
##    0.94        47:3           2.647                -10.667               0.026
##                          (2.079, 13.996)      (-10.667, -10.667)      (-0.109, 0.129)
##
##    0.95        19:1           2.353                 -13                  0.024
##                           (1.596, 13.5)         (-13, -13)            (-0.141, 0.123)
##
##    0.96        24:1           2.059                -16.5                 0.021
##                          (1.348, 13.069)       (-16.5, -16.5)         (-0.132, 0.119)
##
##    0.97        97:3           1.765                -22.333               0.018
##                          (0.928, 11.887)      (-22.333, -22.333)      (-0.189, 0.112)
##
##    0.98        49:1           1.471                 -34                  0.015
##                          (0.545, 10.606)        (-34, -34)            (-0.29, 0.101)
##
##    0.99        99:1           0.882                 -69                  0.009
##                          (0.278, 8.077)         (-69, -69)              (0, 0.08)
```

30

```
##
##        1            Inf:1             0                NA                    NA
##                                    (0, 0)          (NA, NA)              (NA, NA)
## --------------------------------------------------------------------------------
```

```r
plot_clinical_impact(complex, population.size = 1000, cost.benefit.axis = T,
                     n.cost.benefits = 8, col = c('red', 'blue'),
                     confidence.intervals = T,
                     ylim = c(0, 1000),
                     legend.position = 'topright')
```
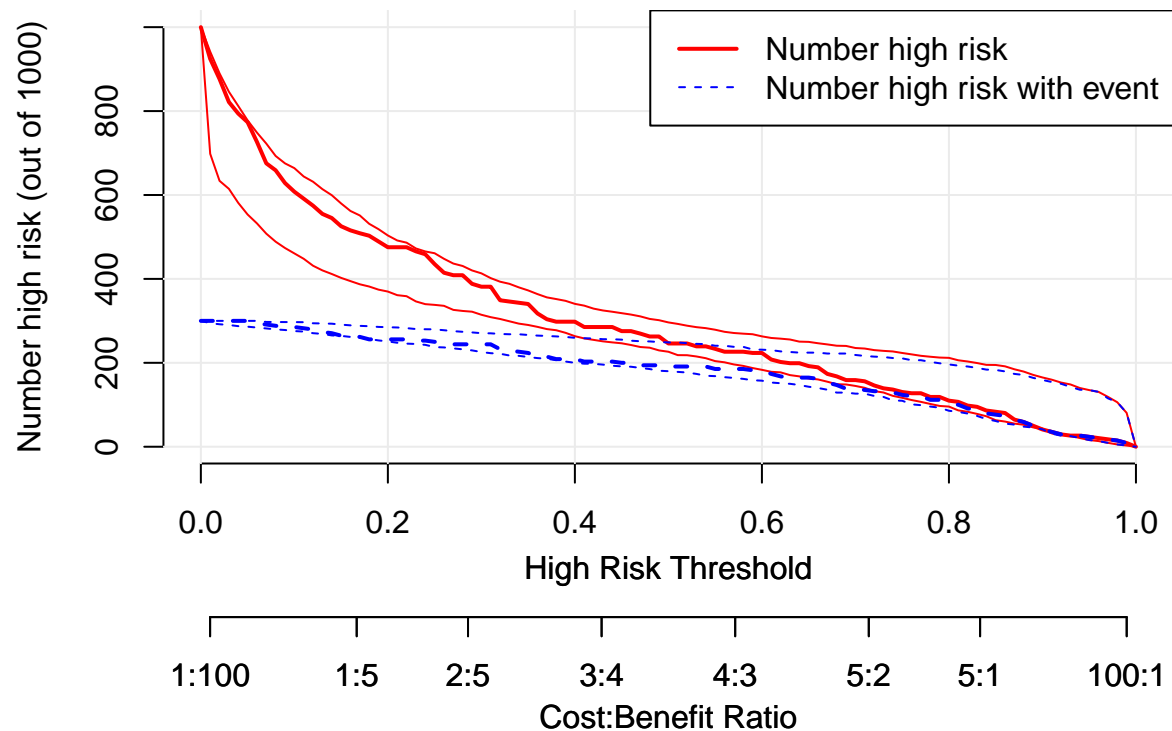

The red curve (number of high risk) indicates the number of people classified as positive (high risk) by the complex model at each threshold probability; the blue curve [(number high) risk with outcome] is the number of true positives for each threshold probability.

```r
mod <- glm(Class2 ~ Symptoms + Endemic + PVT + Metastasis + Age + PS +
           Encephalopathy + Ascites + INR + AFP + Hemoglobin + Leucocytes +
           Albumin + Total_Bil + AST + GGT + ALP + Creatinine + Major_Dim +
           Dir_Bil + Ferritin, data=dt, family=binomial)

h1 <- hoslem.test(mod$y, fitted(mod), g=10)
h1
```

External validation of Logistic regression model:

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  mod$y, fitted(mod)
## X-squared = 9.4905, df = 8, p-value = 0.3026
```

P value is 0.3026, It cannot be considered that the model does not fit well yet.

```
cbind(h1$observed, h1$expected)
```

```
##                      y0 y1      yhat0       yhat1
## [3.91e-05,0.0541] 21  0 20.4961649  0.5038351
## (0.0541,0.13]     18  2 18.1238172  1.8761828
## (0.13,0.211]      17  3 16.6864269  3.3135731
## (0.211,0.341]     11 10 15.2750427  5.7249573
## (0.341,0.513]     15  5 11.2968977  8.7031023
## (0.513,0.617]      8 12  8.9178814 11.0821186
## (0.617,0.795]      7 14  5.7901046 15.2098954
## (0.795,0.888]      4 16  3.1244734 16.8755266
## (0.888,0.942]      1 19  1.6425664 18.3574336
## (0.942,0.999]      0 21  0.6466249 20.3533751
```

Generate Hosmer-Lemeshow contingency table. Among them, y0 is the number of events that didn't occur; y1 is the number of events that occurred; yhat0 represents the probability of events that will not happen predicted by the model; yhat1 represents the probability of events that will happen predicted by the model.

```
pre <- predict(mod, dt, type = c("response"))

h2 <- hoslem.test(dt$Class2, pre, g=10)
h2
```
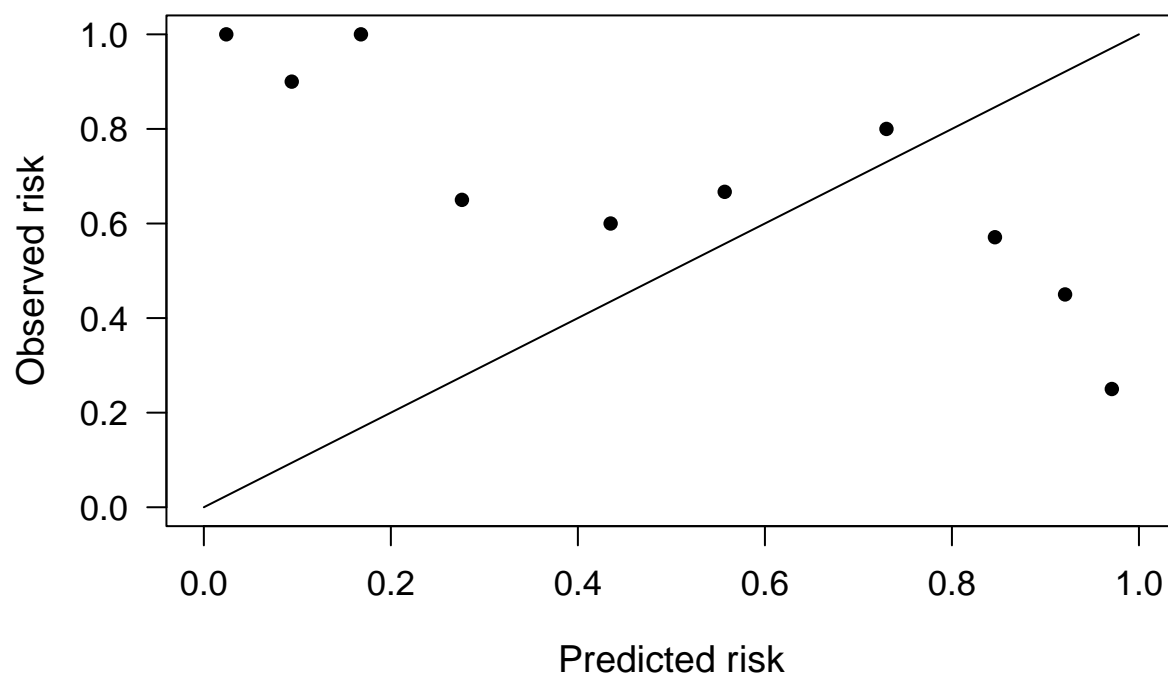
```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  dt$Class2, pre
## X-squared = 9.4905, df = 8, p-value = 0.3026
```

P=0.3026 > 0.05. So, it cannot be considered that the model fitting is poor, suggesting that the model performs well in the new data set. If P < 0.05, the model is poorly fitted.

```
library(PredictABEL)
plotCalibration(data = dt,
                cOutcome = 2,
                predRisk = pre,
                rangeaxis = c(0, 1))
```
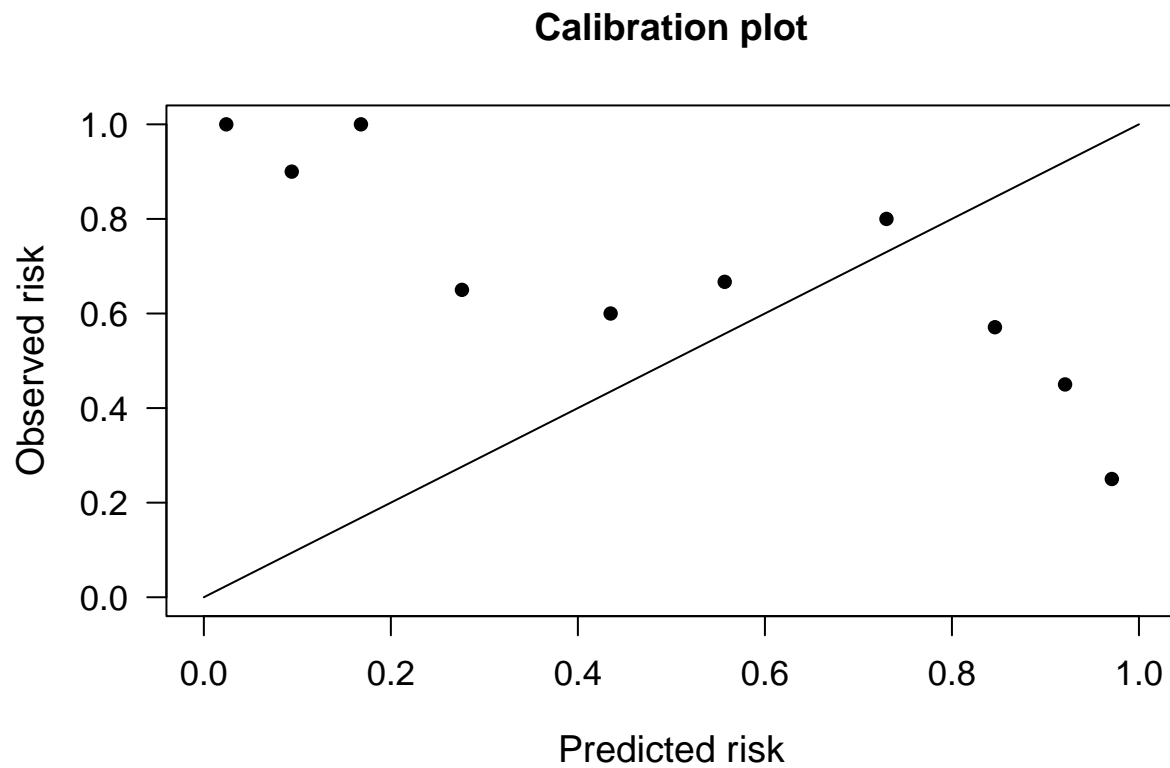
Calibration plot


```
## $Table_HLtest
##                   total meanpred meanobs predicted observed
## [3.91e-05,0.0587)    21    0.024   1.000      0.50       21
## [5.87e-02,0.1319)    20    0.094   0.900      1.88       18
## [1.32e-01,0.2190)    21    0.168   1.000      3.53       21
## [2.19e-01,0.3462)    20    0.276   0.650      5.51       13
## [3.46e-01,0.5135)    20    0.435   0.600      8.70       12
## [5.13e-01,0.6473)    21    0.557   0.667     11.70       14
## [6.47e-01,0.8083)    20    0.730   0.800     14.59       16
## [8.08e-01,0.8902)    21    0.846   0.571     17.76       12
## [8.90e-01,0.9453)    20    0.921   0.450     18.41        9
## [9.45e-01,0.9993]    20    0.971   0.250     19.41        5
##
## $Chi_square
## [1] 1570.332
##
## $df
## [1] 8
##
## $p_value
## [1] 0
```

```
library(PredictABEL)
library(pROC)
plotCalibration(data = dt,
```

```
              cOutcome = 2,
              predRisk = pre,
              rangeaxis = c(0, 1))
```

Calibration plot


```
## $Table_HLtest
##                      total meanpred meanobs predicted observed
## [3.91e-05,0.0587)      21    0.024   1.000      0.50       21
## [5.87e-02,0.1319)      20    0.094   0.900      1.88       18
## [1.32e-01,0.2190)      21    0.168   1.000      3.53       21
## [2.19e-01,0.3462)      20    0.276   0.650      5.51       13
## [3.46e-01,0.5135)      20    0.435   0.600      8.70       12
## [5.13e-01,0.6473)      21    0.557   0.667     11.70       14
## [6.47e-01,0.8083)      20    0.730   0.800     14.59       16
## [8.08e-01,0.8902)      21    0.846   0.571     17.76       12
## [8.90e-01,0.9453)      20    0.921   0.450     18.41        9
## [9.45e-01,0.9993]      20    0.971   0.250     19.41        5
##
## $Chi_square
## [1] 1570.332
##
## $df
## [1] 8
##
## $p_value
## [1] 0
```

```
library(pROC)
pr <- predict(mod, type = c("response"))

roc_curve <- pROC::roc(dt$Class2 ~ pr)


auc(roc_curve)
```
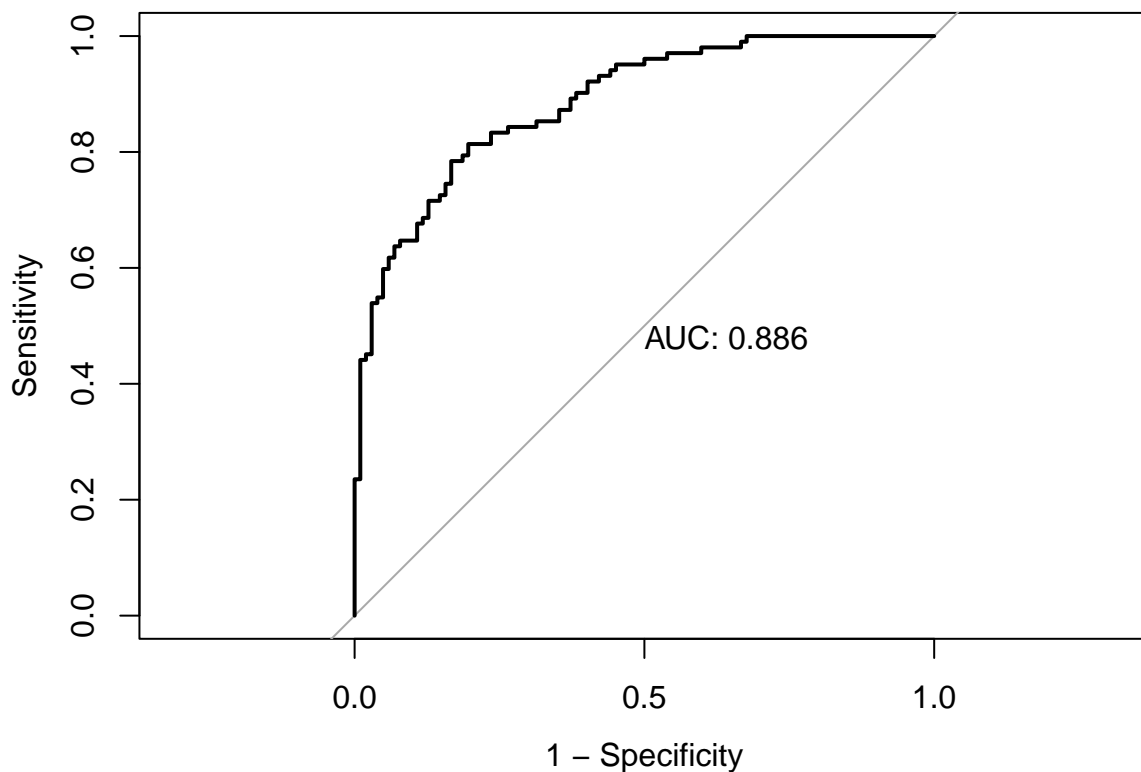
Discrimination evaluation:

```
## Area under the curve: 0.8861
```

```
plot(roc_curve,
     print.cutoffs.at=-c(50,90,110,120),
     cutoff.label.function=`-`,
     points.pch = 20, points.col = "darkblue",
     text.adj=c(1.2,-0.5), legacy.axes = TRUE,
     print.auc = TRUE)
```


The model was verified in the external data set, AUC =0.8861, indicating that the model has a good discrimination in the validation of the external data set.