

```
$dataFile
```

```
[1] "payer_model_SY_GP&iOS_mkt_2019-01-01_2019-03-31.rds"
```

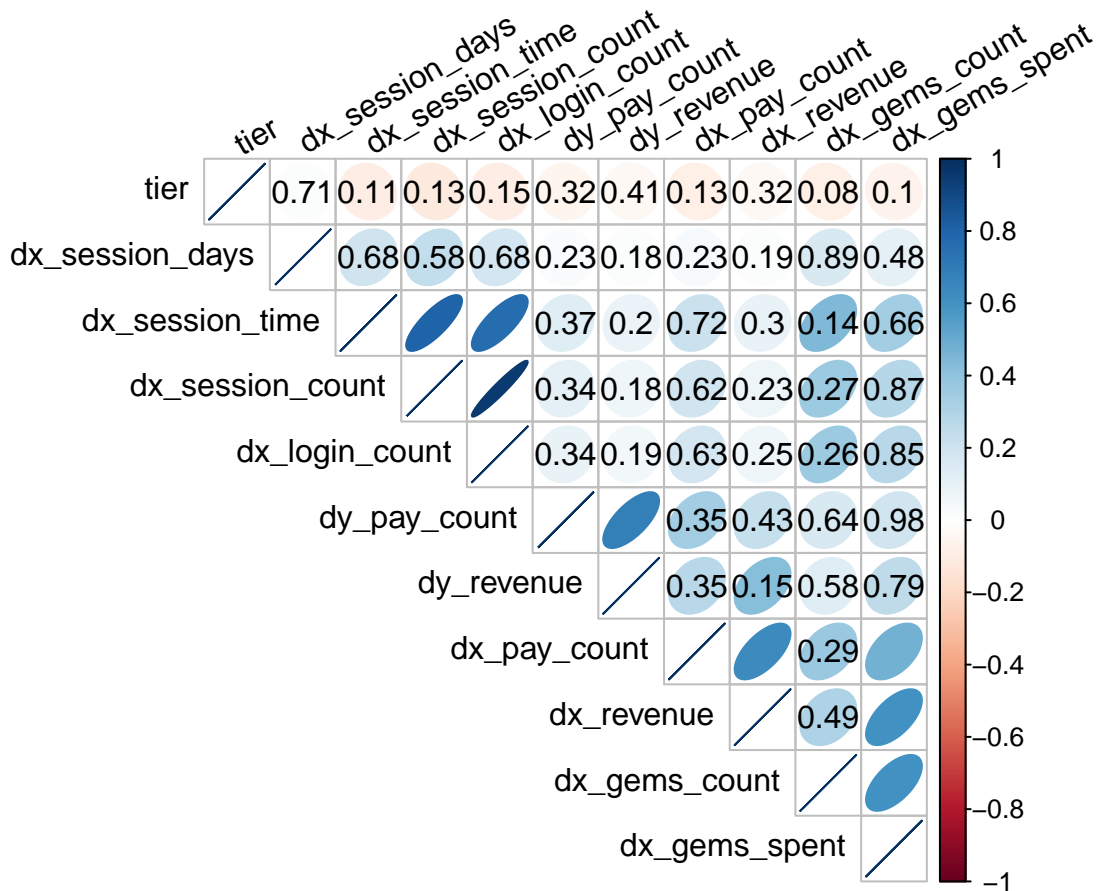
```
$strainRegDate
```

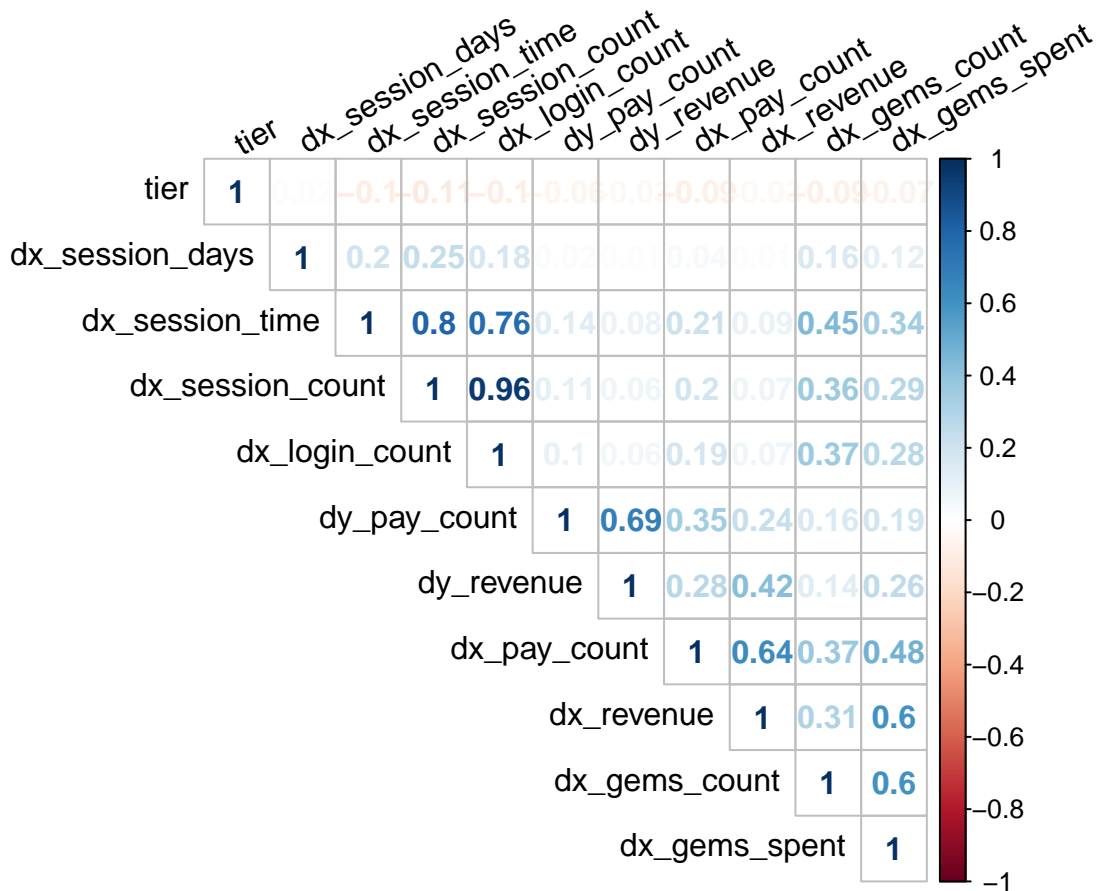
```
[1] "2019-01-01" "2019-01-31"
```

```
$testRegDate
```

```
[1] "2019-03-01" "2019-03-31"
```

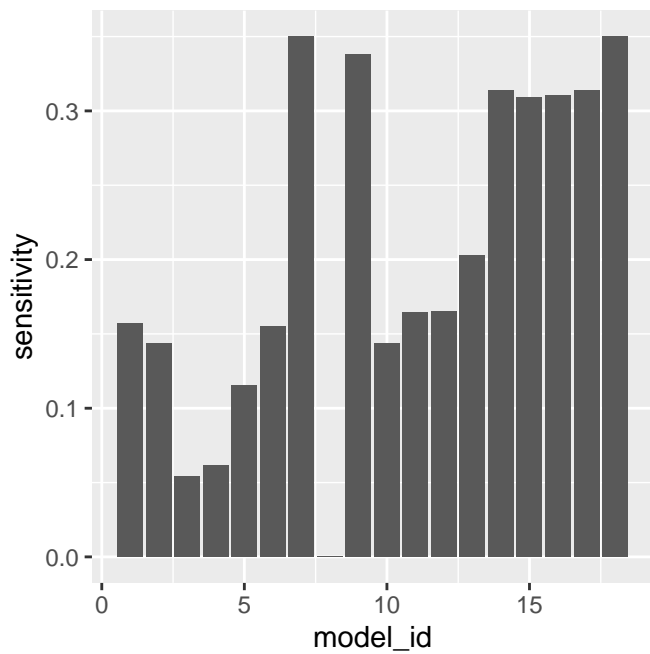
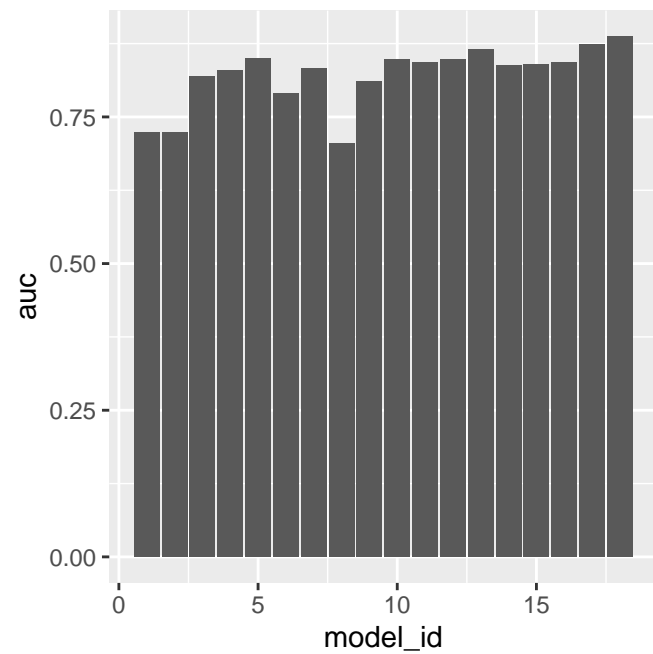
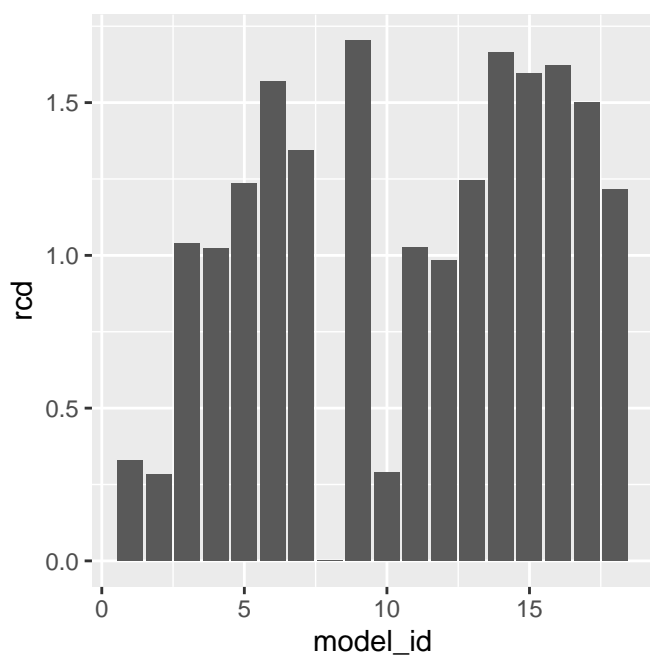
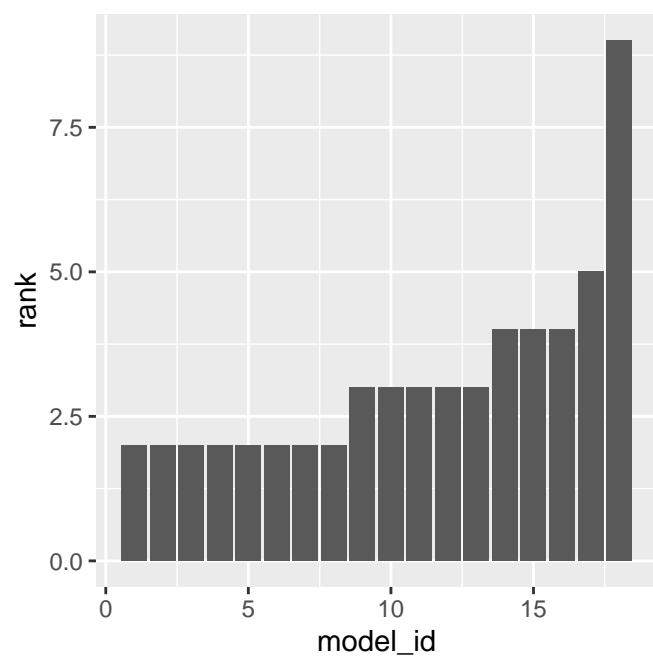
```
[1] "0_data/payer_model_SY_GP&iOS_mkt_2019-01-01_2019-03-31.rds"
```





[illegible]

	model_id	rank	auc	cut_off	rcd	sensitivity
1	1	2	0.7233216	0.01374235	0.3276247	0.15711095
2	2	2	0.7232533	0.38201185	0.2836932	0.14370812
3	3	2	0.8197521	0.06210410	1.0402085	0.05435592
4	4	2	0.8293952	0.06721169	1.0230827	0.06180194
5	5	2	0.8498433	0.06574873	1.2352941	0.11541325
6	6	2	0.7893548	0.05578675	1.5703649	0.15487714
7	7	2	0.8321707	0.04758432	1.3425168	0.34996277
8	8	2	0.7057662	0.99993389	0.0000000	0.00000000
9	9	3	0.8113187	0.07749460	1.7029039	0.33804914
10	10	3	0.8474211	0.23611382	0.2889054	0.14370812
11	11	3	0.8421485	0.06584952	1.0245719	0.16455696
12	12	3	0.8484114	0.06914338	0.9843634	0.16530156
13	13	3	0.8647064	0.06541980	1.2442293	0.20253165
14	14	4	0.8374836	0.09093946	1.6656739	0.31347729
15	15	4	0.8401108	0.09350873	1.5971705	0.30900968
16	16	4	0.8428292	0.09556242	1.6209978	0.31049888
17	17	5	0.8731825	0.10519414	1.5018615	0.31347729
18	18	9	0.8868971	0.13424567	1.2166791	0.34996277



THIS SUMMARY IS OUTDATED

WE FOUND OUT THAT WE NEED CROSS-VALIDATION TO GET ROBUST RESULTS

Summary for SY:

Based on correlation matrix, these variables should not be together

dx_session_time, dx_session_count, dx_login_count

dx_pay_count, dx_revenue

dx_gems_spent, dx_gems_count

dx_gems_spent, dx_revenue

dx_gems_spent, dx_pay_count

We will test 1-variable models

dx_revenue ~ dx_pay_count (we choose revenue)

dx_session_time > dx_session_count > dx_login_count (as expected)

dx_gems_spent > dx_gems_count

Check session features with dx_revenue

dx_session_time > dx_session_count > dx_login_count (as expected)

Check gem features with previous

dx_gems_spent > dx_gems_count

even though gems_spent is correlated with revenue it makes a better model

Check tier

tier barely improve the model

The full model is significantly better compared to reduced model.

It should be compared with cross-validation whether this difference is robust.

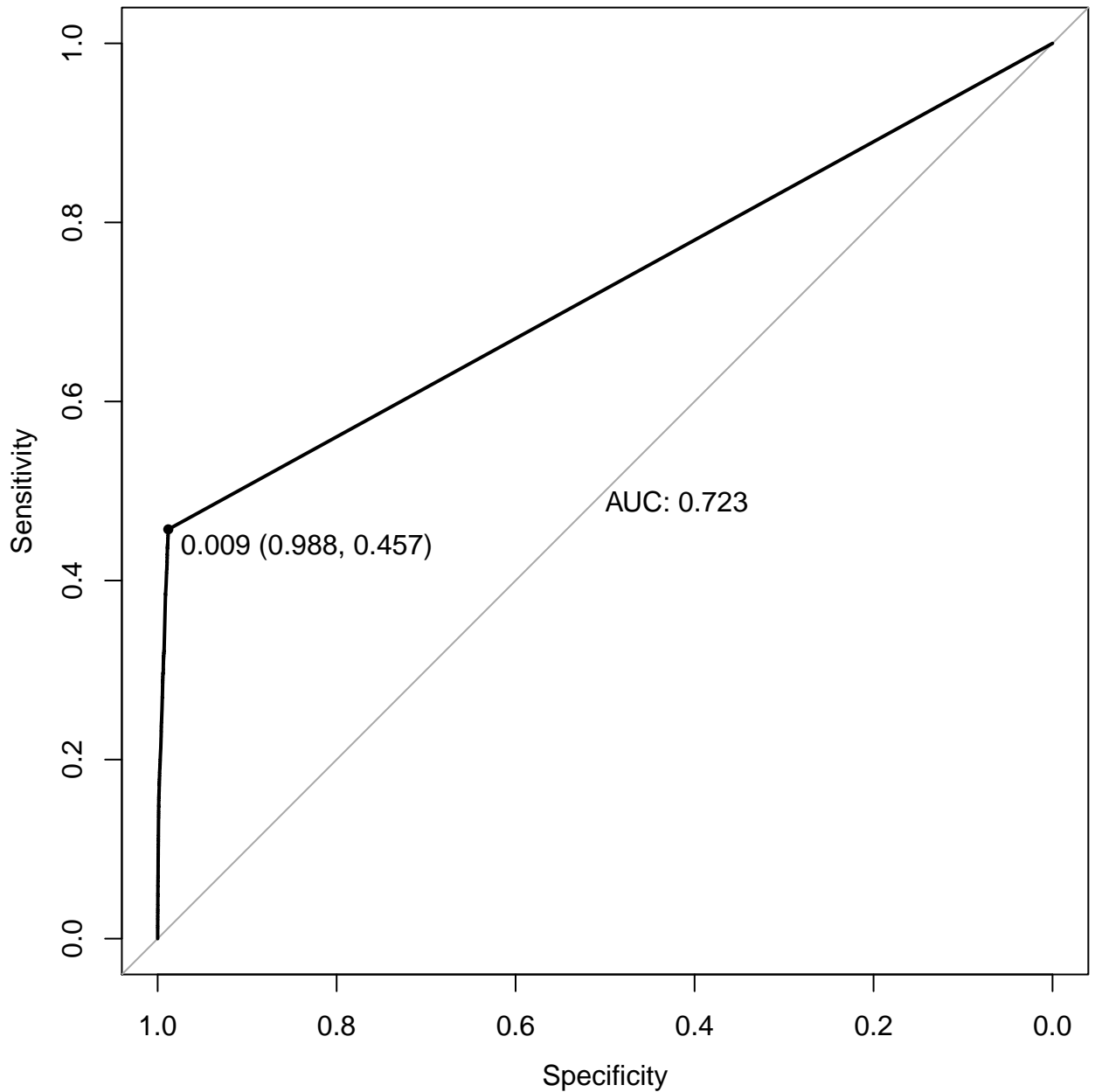
Best model so far has all variables.

Maybe removing the features from the full model makes sense.

Multicollinearity is present in the full model but is it a problem?

dx_gem_spent is suspiciously good, is it robust? We need CV...

ROC curve for model 1



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1167  -0.1167  -0.1167   3.1601

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.986315   0.020521 -242.99  <2e-16 ***
dx_revenue   0.245079   0.008185   29.94  <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: 30587  on 350797  degrees of freedom
Residual deviance: 28973  on 350796  degrees of freedom
AIC: 28977

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.7233
```

```
$relativeCountDifference
```

```
[1] 0.3276247
```

```
$optimal_cut_off
```

```
[1] 0.01374235
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	174407	1132
TRUE	229	211

Accuracy : 0.9923

95% CI : (0.9918, 0.9927)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 0.695

Kappa : 0.2338

McNemar's Test P-Value : <2e-16

Sensitivity : 0.157111

Specificity : 0.998689

Pos Pred Value : 0.479545

Neg Pred Value : 0.993551

Prevalence : 0.007632

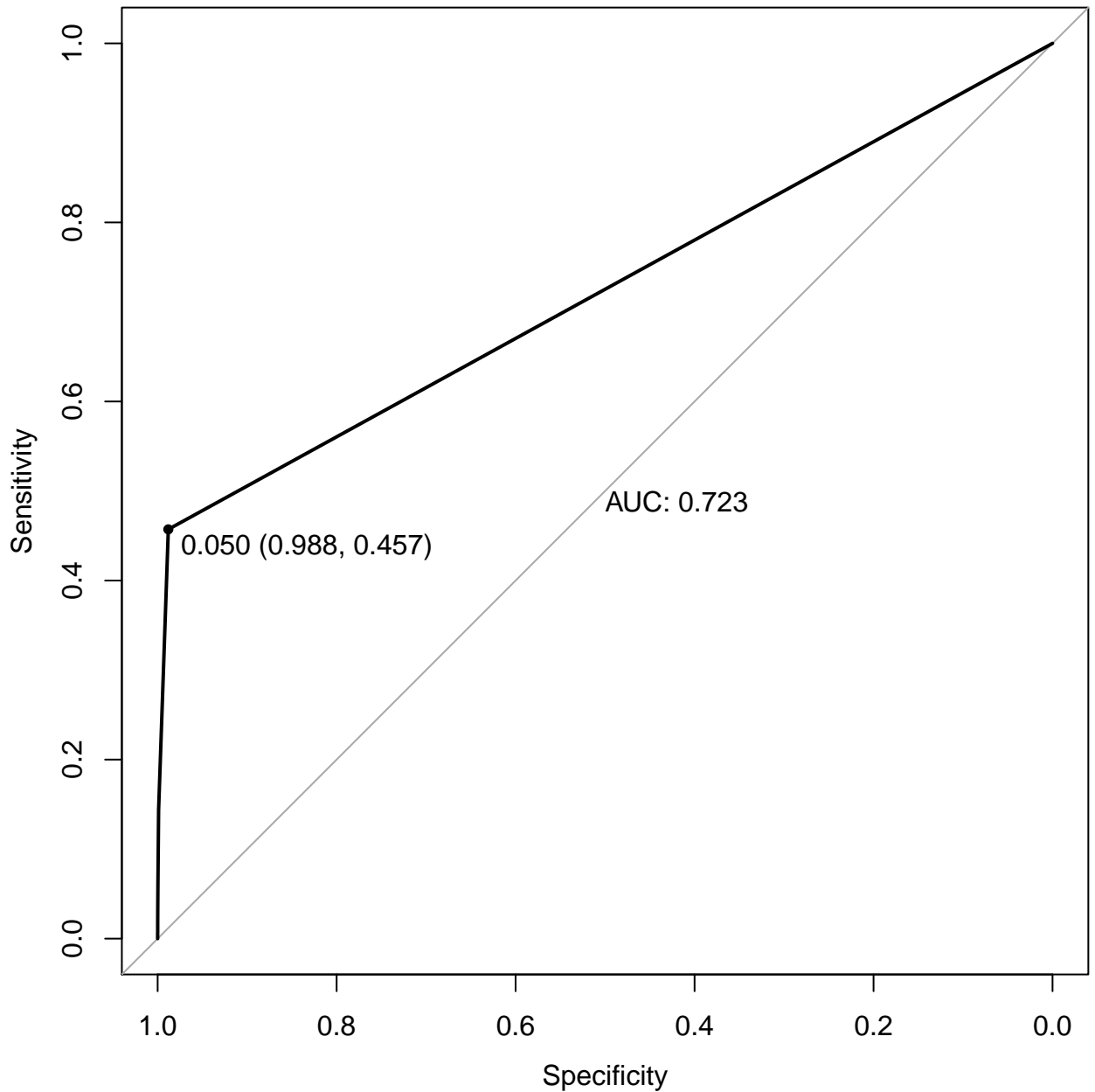
Detection Rate : 0.001199

Detection Prevalence : 0.002500

Balanced Accuracy : 0.577900

'Positive' Class : TRUE

ROC curve for model 2



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.0986  -0.0986  -0.0986   3.2644

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.32323     0.02413  -220.61  <2e-16 ***
dx_pay_count   3.07168     0.04095   75.01  <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: 30587  on 350797  degrees of freedom
Residual deviance: 24669  on 350796  degrees of freedom
AIC: 24673

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.7233
```

```
$relativeCountDifference
```

```
[1] 0.2836932
```



```
$optimal_cut_off
```

```
[1] 0.3820119
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	174448	1150
TRUE	188	193

Accuracy : 0.9924

95% CI : (0.992, 0.9928)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 0.4527

Kappa : 0.2213

McNemar's Test P-Value : <2e-16

Sensitivity : 0.143708

Specificity : 0.998923

Pos Pred Value : 0.506562

Neg Pred Value : 0.993451

Prevalence : 0.007632

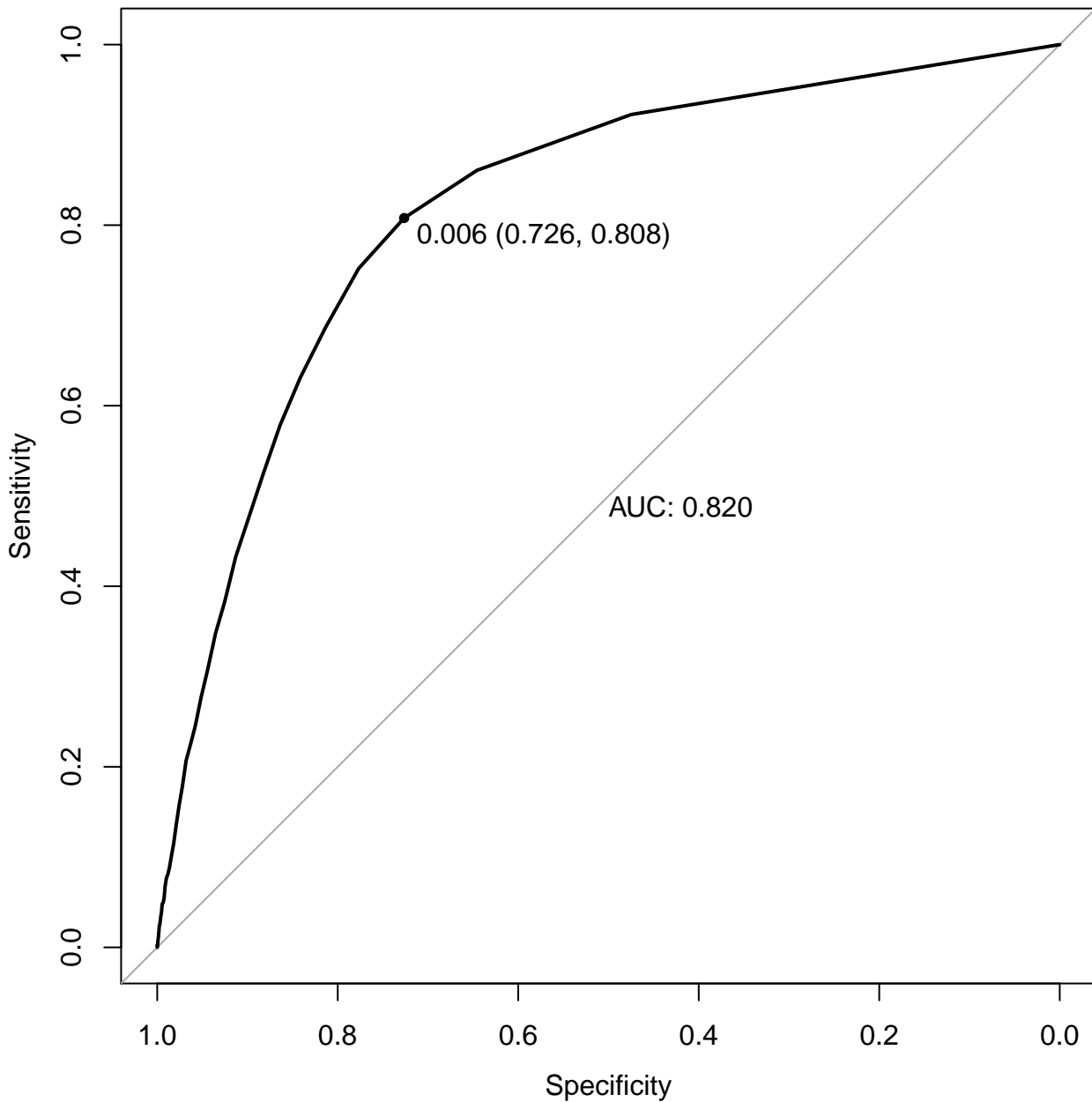
Detection Rate : 0.001097

Detection Prevalence : 0.002165

Balanced Accuracy : 0.571316

'Positive' Class : TRUE

ROC curve for model 3



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0001  -0.1043  -0.0941  -0.0941   3.3241

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.520806   0.026204 -210.68  <2e-16 ***
dx_login_count  0.103194   0.001713   60.23  <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: 30587  on 350797  degrees of freedom
Residual deviance: 28141  on 350796  degrees of freedom
AIC: 28145

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8198
```

```
$relativeCountDifference
```

```
[1] 1.040208
```

```
$optimal_cut_off
```

```
[1] 0.0621041
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173312	1270
TRUE	1324	73

Accuracy : 0.9853
95% CI : (0.9847, 0.9858)
No Information Rate : 0.9924
P-Value [Acc > NIR] : 1.0000

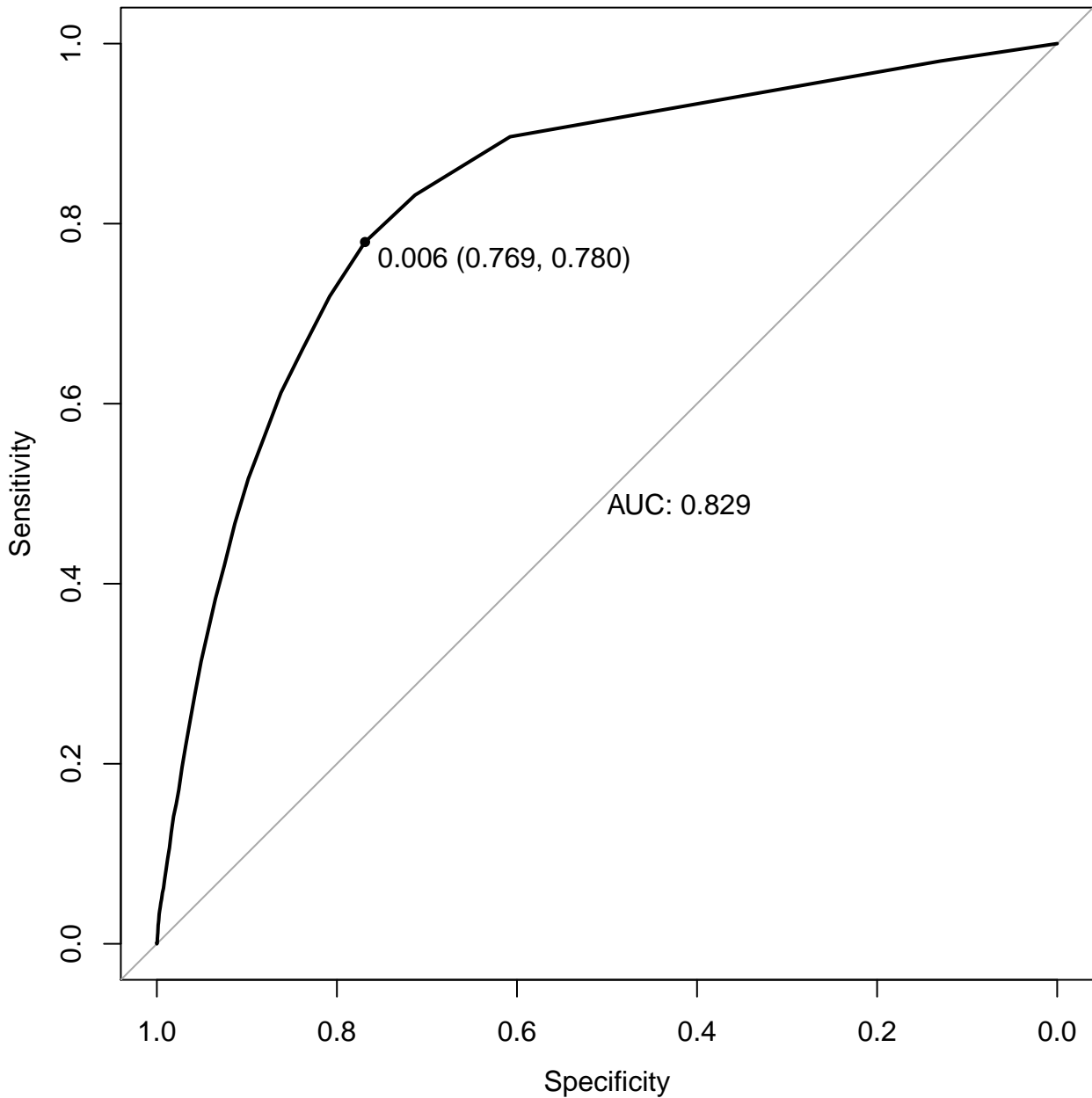
Kappa : 0.0459

McNemar's Test P-Value : 0.2981

Sensitivity : 0.0543559
Specificity : 0.9924185
Pos Pred Value : 0.0522548
Neg Pred Value : 0.9927255
Prevalence : 0.0076316
Detection Rate : 0.0004148
Detection Prevalence : 0.0079384
Balanced Accuracy : 0.5233872

'Positive' Class : TRUE

ROC curve for model 4




```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7562  -0.1051  -0.0942  -0.0942   3.3257

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.526288   0.026194  -211.0  <2e-16 ***
dx_session_count 0.110124   0.001715   64.2  <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: 30587  on 350797  degrees of freedom
Residual deviance: 27797  on 350796  degrees of freedom
AIC: 27801

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8294
```

```
$relativeCountDifference
```

```
[1] 1.023083
```

```
$optimal_cut_off
```

```
[1] 0.06721169
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173345	1260
TRUE	1291	83

Accuracy : 0.9855

95% CI : (0.9849, 0.9861)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1.0000

Kappa : 0.0538

McNemar's Test P-Value : 0.5525

Sensitivity : 0.0618019

Specificity : 0.9926075

Pos Pred Value : 0.0604076

Neg Pred Value : 0.9927837

Prevalence : 0.0076316

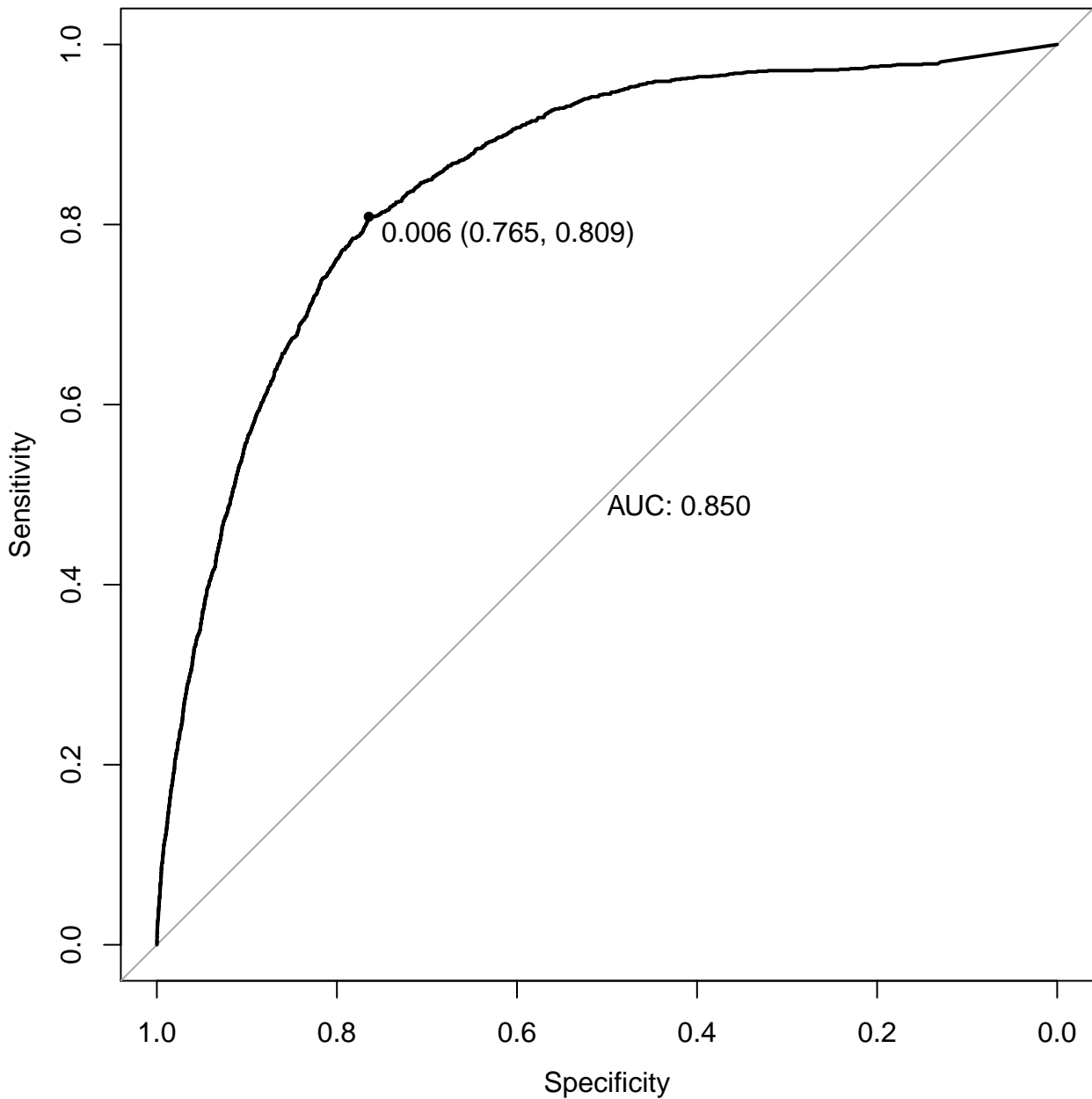
Detection Rate : 0.0004716

Detection Prevalence : 0.0078077

Balanced Accuracy : 0.5272047

'Positive' Class : TRUE

ROC curve for model 5



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8701  -0.1056  -0.0939  -0.0921   3.3090

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.4703801  0.0252398 -216.74  <2e-16 ***
dx_session_time  0.0073346  0.0001077   68.13  <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: 30587  on 350797  degrees of freedom
Residual deviance: 27376  on 350796  degrees of freedom
AIC: 27380

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8498
```

```
$relativeCountDifference
```

```
[1] 1.235294
```

```
$optimal_cut_off
```

```
[1] 0.06574873
```


\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173132	1188
TRUE	1504	155

Accuracy : 0.9847

95% CI : (0.9841, 0.9853)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.0956

McNemar's Test P-Value : 1.27e-09

Sensitivity : 0.1154133

Specificity : 0.9913878

Pos Pred Value : 0.0934298

Neg Pred Value : 0.9931849

Prevalence : 0.0076316

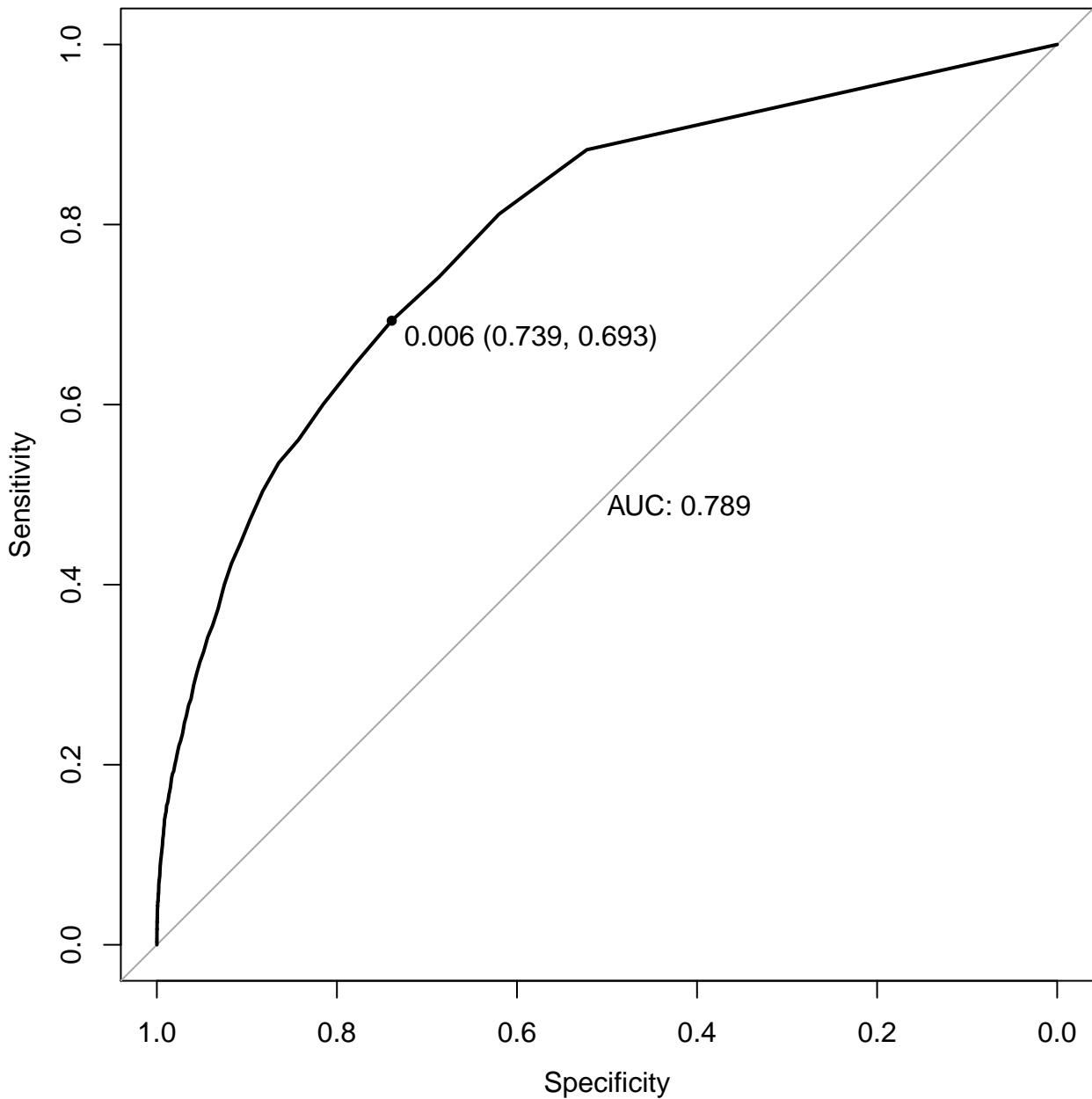
Detection Rate : 0.0008808

Detection Prevalence : 0.0094273

Balanced Accuracy : 0.5534005

'Positive' Class : TRUE

ROC curve for model 6



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.8265  -0.1069  -0.0969  -0.0969   3.2752

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.3586924  0.0239765 -223.50  <2e-16 ***
dx_gems_count  0.0654927  0.0009934   65.93  <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: 30587  on 350797  degrees of freedom
Residual deviance: 27303  on 350796  degrees of freedom
AIC: 27307

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.7894
```

```
$relativeCountDifference
```

```
[1] 1.570365
```

```
$optimal_cut_off
```

```
[1] 0.05578675
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	172735	1135
TRUE	1901	208

Accuracy : 0.9827
95% CI : (0.9821, 0.9834)
No Information Rate : 0.9924
P-Value [Acc > NIR] : 1

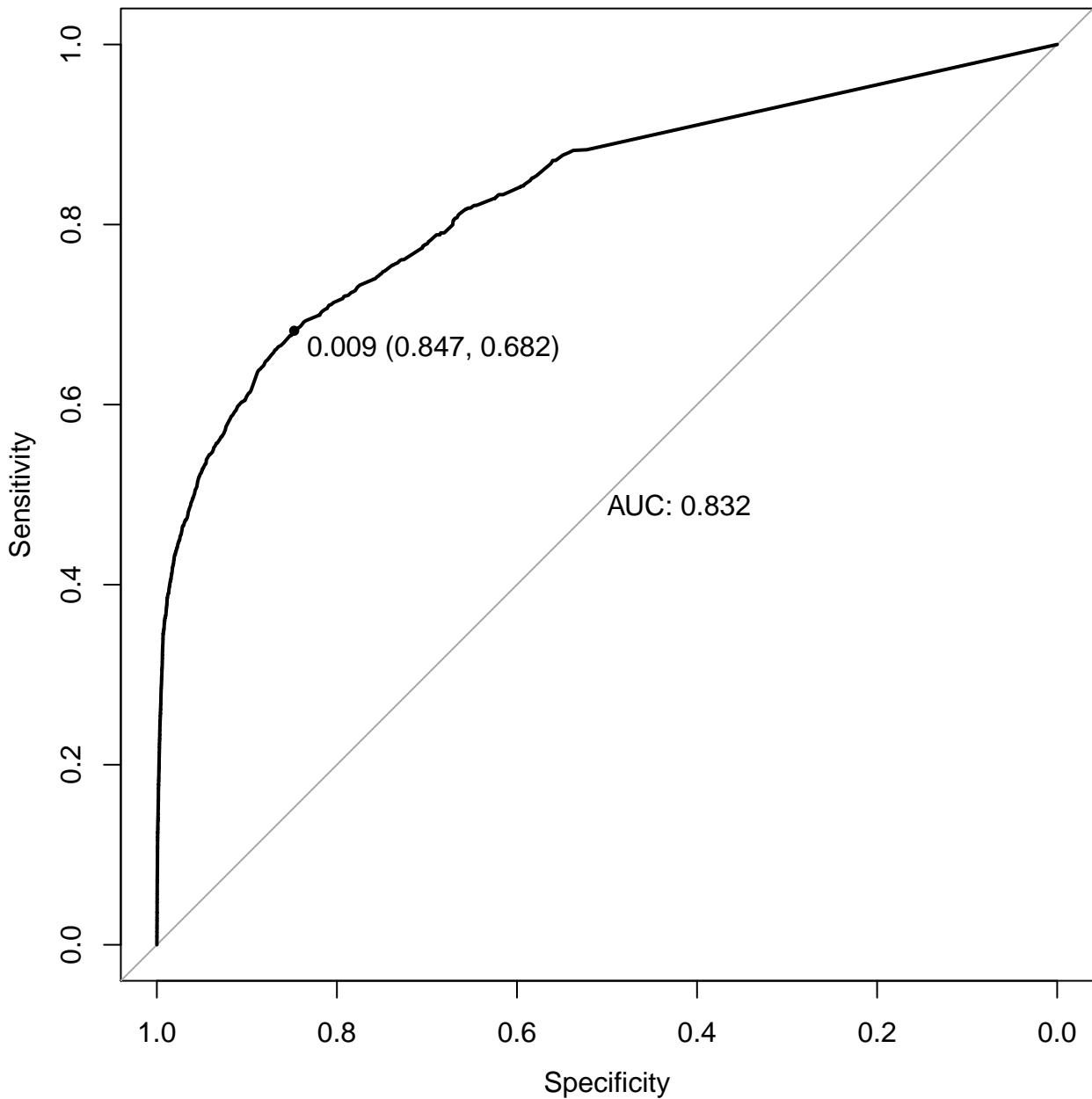
Kappa : 0.1122

McNemar's Test P-Value : <2e-16

Sensitivity : 0.154877
Specificity : 0.989115
Pos Pred Value : 0.098625
Neg Pred Value : 0.993472
Prevalence : 0.007632
Detection Rate : 0.001182
Detection Prevalence : 0.011984
Balanced Accuracy : 0.571996

'Positive' Class : TRUE

ROC curve for model 7



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1162  -0.0895  -0.0895   3.3233

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.5180696  0.0256508 -215.12  <2e-16 ***
dx_gems_spent  0.0065453  0.0001074   60.95  <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: 30587  on 350797  degrees of freedom
Residual deviance: 25830  on 350796  degrees of freedom
AIC: 25834

Number of Fisher Scoring iterations: 9

```



```
$auc
```

```
Area under the curve: 0.8322
```

```
$relativeCountDifference
```

```
[1] 1.342517
```

```
$optimal_cut_off
```

```
[1] 0.04758432
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173303	873
TRUE	1333	470

Accuracy : 0.9875

95% CI : (0.9869, 0.988)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.2926

McNemar's Test P-Value : <2e-16

Sensitivity : 0.349963

Specificity : 0.992367

Pos Pred Value : 0.260677

Neg Pred Value : 0.994988

Prevalence : 0.007632

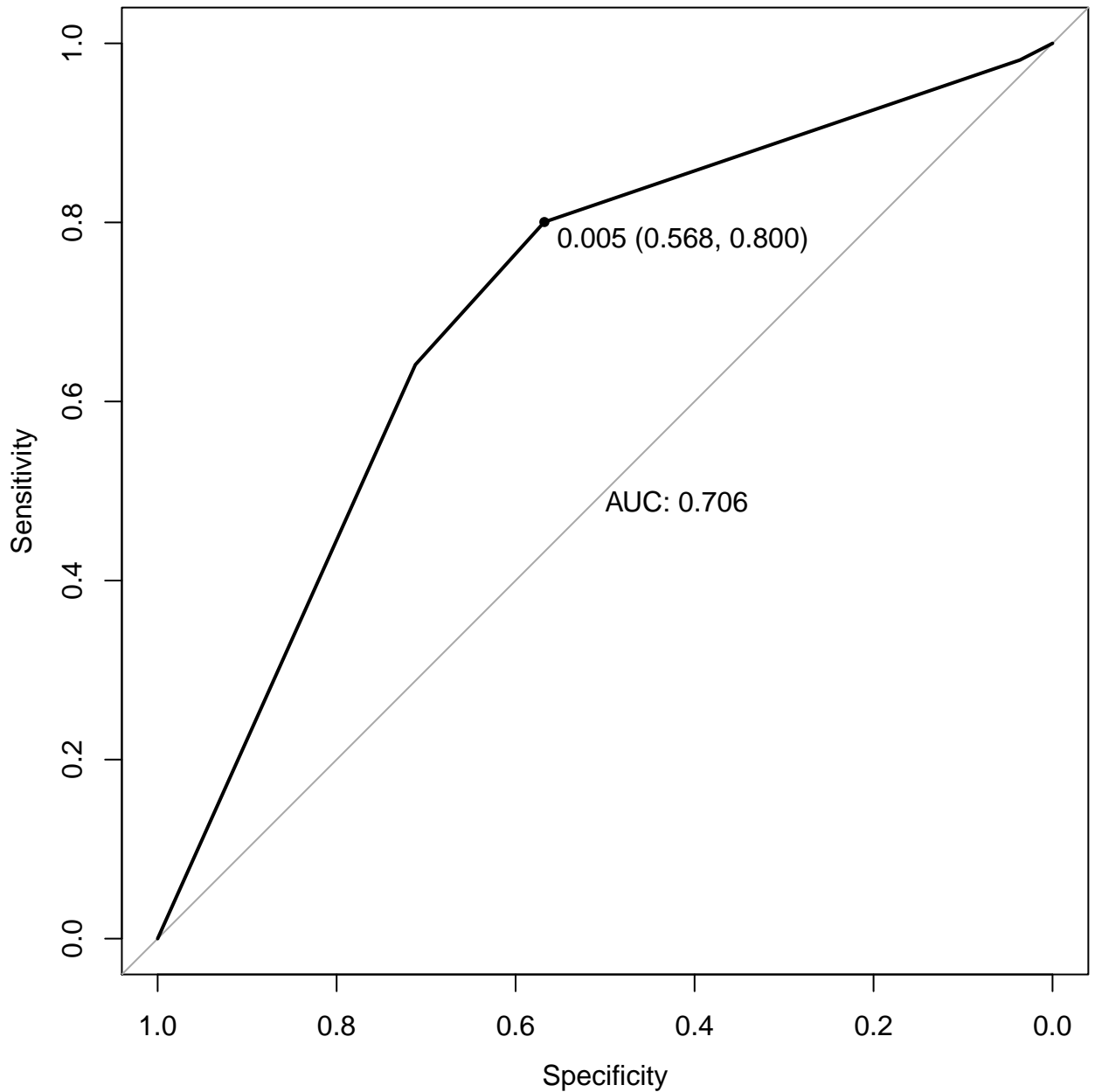
Detection Rate : 0.002671

Detection Prevalence : 0.010246

Balanced Accuracy : 0.671165

'Positive' Class : TRUE

ROC curve for model 8



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.1884  -0.1884  -0.0737  -0.0737   3.7014

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.08090     0.04253  -72.43  <2e-16 ***
tier         -0.94209     0.02411  -39.08  <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: 30587  on 350797  degrees of freedom
Residual deviance: 28770  on 350796  degrees of freedom
AIC: 28774

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.7058
```

```
$relativeCountDifference
```

```
[1] 0
```

```
$optimal_cut_off
```

```
[1] 0.9999339
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	174636	1343
TRUE	0	0

Accuracy : 0.9924

95% CI : (0.992, 0.9928)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 0.5073

Kappa : 0

McNemar's Test P-Value : <2e-16

Sensitivity : 0.000000

Specificity : 1.000000

Pos Pred Value : NaN

Neg Pred Value : 0.992368

Prevalence : 0.007632

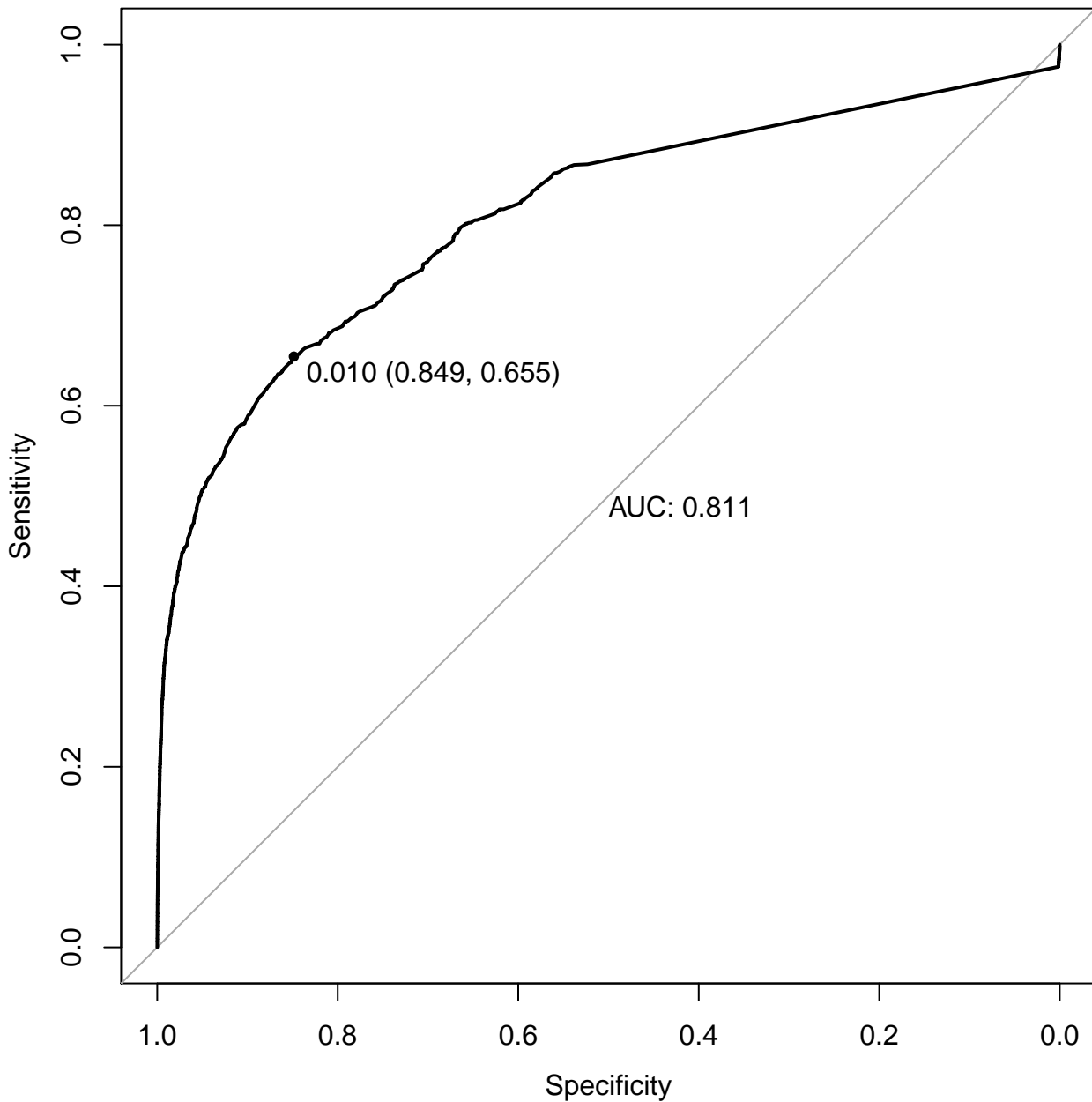
Detection Rate : 0.000000

Detection Prevalence : 0.000000

Balanced Accuracy : 0.500000

'Positive' Class : TRUE

ROC curve for model 9



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1129  -0.0752  -0.0752   5.8163

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.8658942  0.0301834 -194.34  <2e-16 ***
dx_revenue    -0.3998152  0.0117235  -34.10  <2e-16 ***
dx_gems_spent  0.0101719  0.0001493   68.12  <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: 30587  on 350797  degrees of freedom
Residual deviance: 24998  on 350795  degrees of freedom
AIC: 25004

Number of Fisher Scoring iterations: 11

```

```
$auc
```

```
Area under the curve: 0.8113
```

```
$relativeCountDifference
```

```
[1] 1.702904
```

```
$optimal_cut_off
```

```
[1] 0.0774946
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	172803	889
TRUE	1833	454

Accuracy : 0.9845

95% CI : (0.9839, 0.9851)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.2429

McNemar's Test P-Value : <2e-16

Sensitivity : 0.338049

Specificity : 0.989504

Pos Pred Value : 0.198513

Neg Pred Value : 0.994882

Prevalence : 0.007632

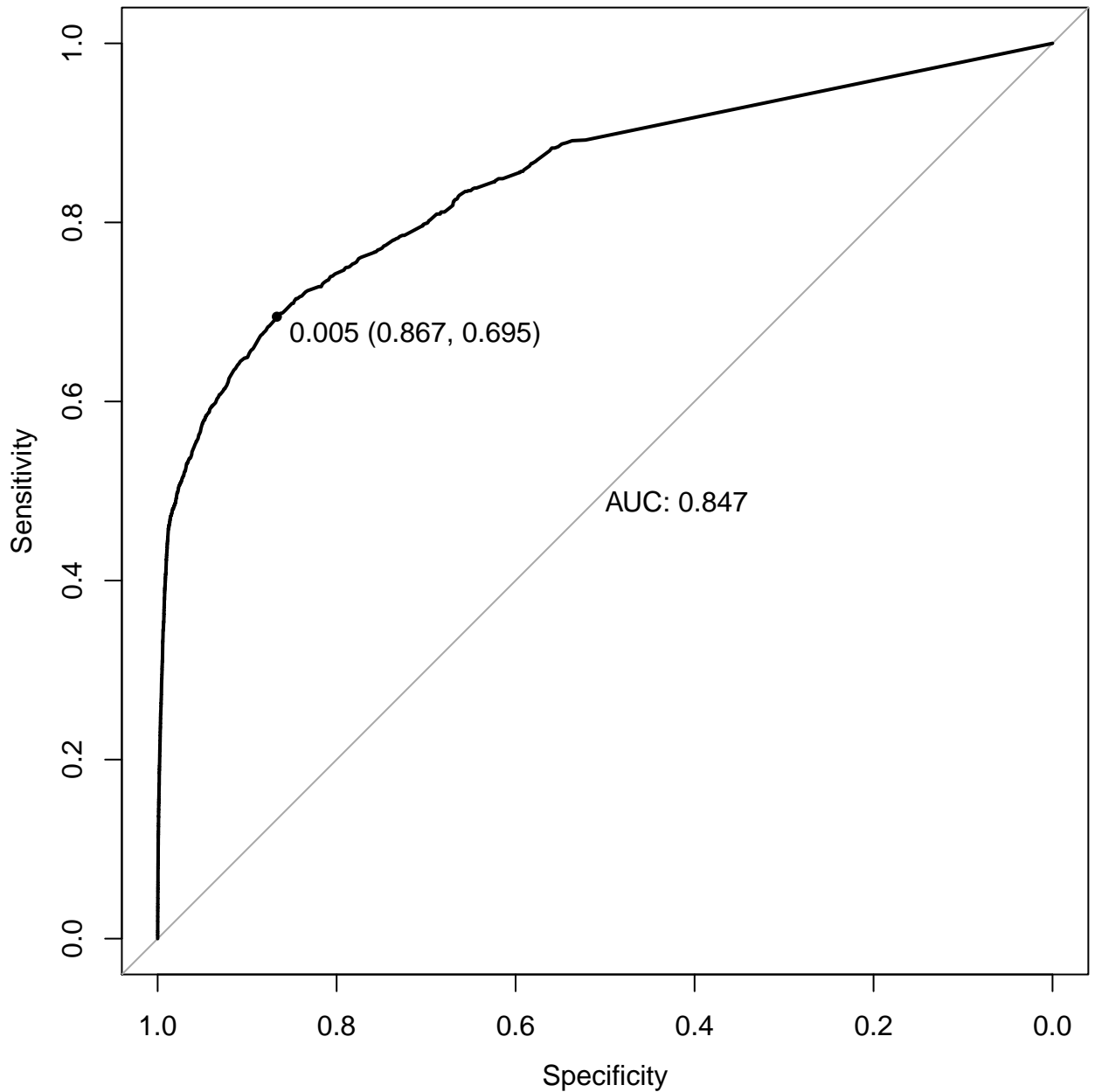
Detection Rate : 0.002580

Detection Prevalence : 0.012996

Balanced Accuracy : 0.663777

'Positive' Class : TRUE

ROC curve for model 10



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1000  -0.0977  -0.0977   3.2703

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.3426792  0.0247038 -216.27 < 2e-16 ***
dx_pay_count   2.8741548  0.0647704  44.38 < 2e-16 ***
dx_gems_spent  0.0005821  0.0001492   3.90 9.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 30587  on 350797  degrees of freedom
Residual deviance: 24646  on 350795  degrees of freedom
AIC: 24652

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8474
```

```
$relativeCountDifference
```

```
[1] 0.2889054
```



```
$optimal_cut_off
```

```
[1] 0.2361138
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	174441	1150
TRUE	195	193

Accuracy : 0.9924

95% CI : (0.9919, 0.9928)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 0.5291

Kappa : 0.2203

McNemar's Test P-Value : <2e-16

Sensitivity : 0.143708

Specificity : 0.998883

Pos Pred Value : 0.497423

Neg Pred Value : 0.993451

Prevalence : 0.007632

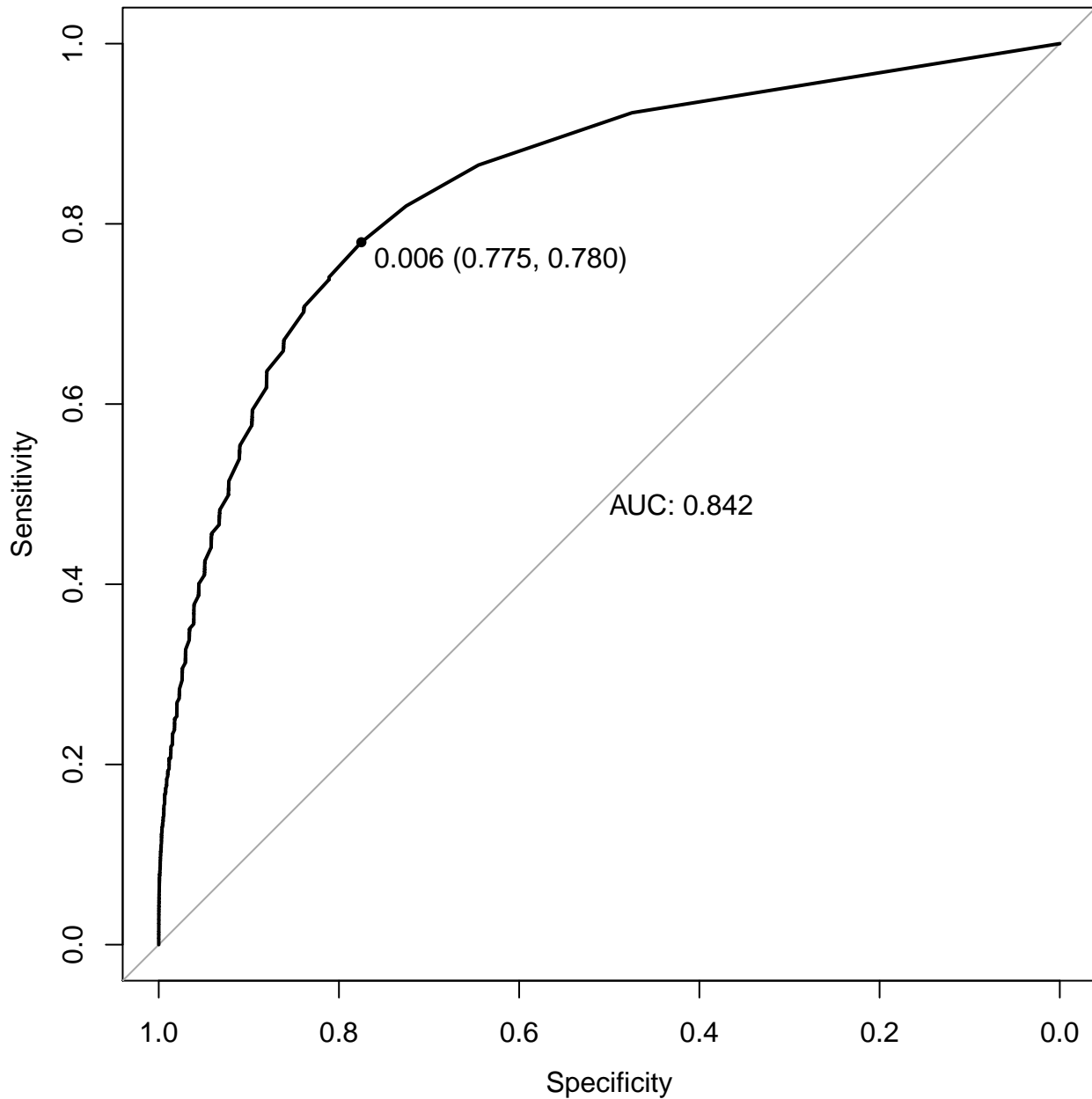
Detection Rate : 0.001097

Detection Prevalence : 0.002205

Balanced Accuracy : 0.571296

'Positive' Class : TRUE

ROC curve for model 11



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1008  -0.0912  -0.0912   3.3415

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.579167   0.027005 -206.60  <2e-16 ***
dx_revenue     0.202466   0.007451  27.17  <2e-16 ***
dx_login_count  0.099536   0.001775  56.06  <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: 30587  on 350797  degrees of freedom
Residual deviance: 26852  on 350795  degrees of freedom
AIC: 26858

Number of Fisher Scoring iterations: 10

```

```
$auc
```

```
Area under the curve: 0.8421
```

```
$relativeCountDifference
```

```
[1] 1.024572
```

```
$optimal_cut_off
```

```
[1] 0.06584952
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173481	1122
TRUE	1155	221

Accuracy : 0.9871

95% CI : (0.9865, 0.9876)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1.0000

Kappa : 0.156

McNemar's Test P-Value : 0.5025

Sensitivity : 0.164557

Specificity : 0.993386

Pos Pred Value : 0.160610

Neg Pred Value : 0.993574

Prevalence : 0.007632

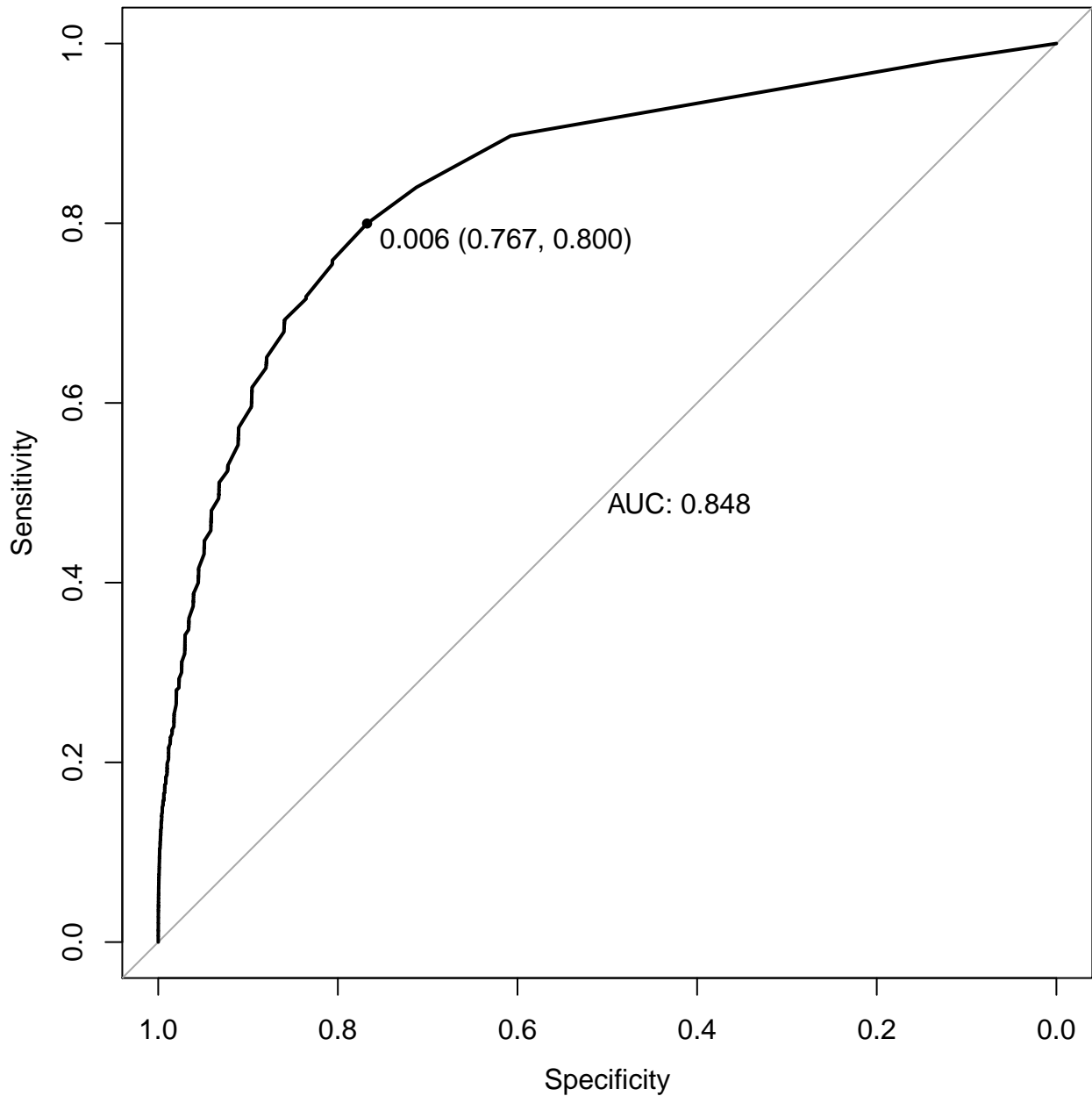
Detection Rate : 0.001256

Detection Prevalence : 0.007819

Balanced Accuracy : 0.578972

'Positive' Class : TRUE

ROC curve for model 12




```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1016  -0.0913  -0.0913   3.3428

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.583554   0.026970  -207.03  <2e-16 ***
dx_revenue     0.195514   0.007339    26.64  <2e-16 ***
dx_session_count 0.106250   0.001777    59.80  <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: 30587  on 350797  degrees of freedom
Residual deviance: 26554  on 350795  degrees of freedom
AIC: 26560

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8484
```

```
$relativeCountDifference
```

```
[1] 0.9843634
```

```
$optimal_cut_off
```

```
[1] 0.06914338
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173536	1121
TRUE	1100	222

Accuracy : 0.9874

95% CI : (0.9868, 0.9879)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1.0000

Kappa : 0.1602

McNemar's Test P-Value : 0.6713

Sensitivity : 0.165302

Specificity : 0.993701

Pos Pred Value : 0.167927

Neg Pred Value : 0.993582

Prevalence : 0.007632

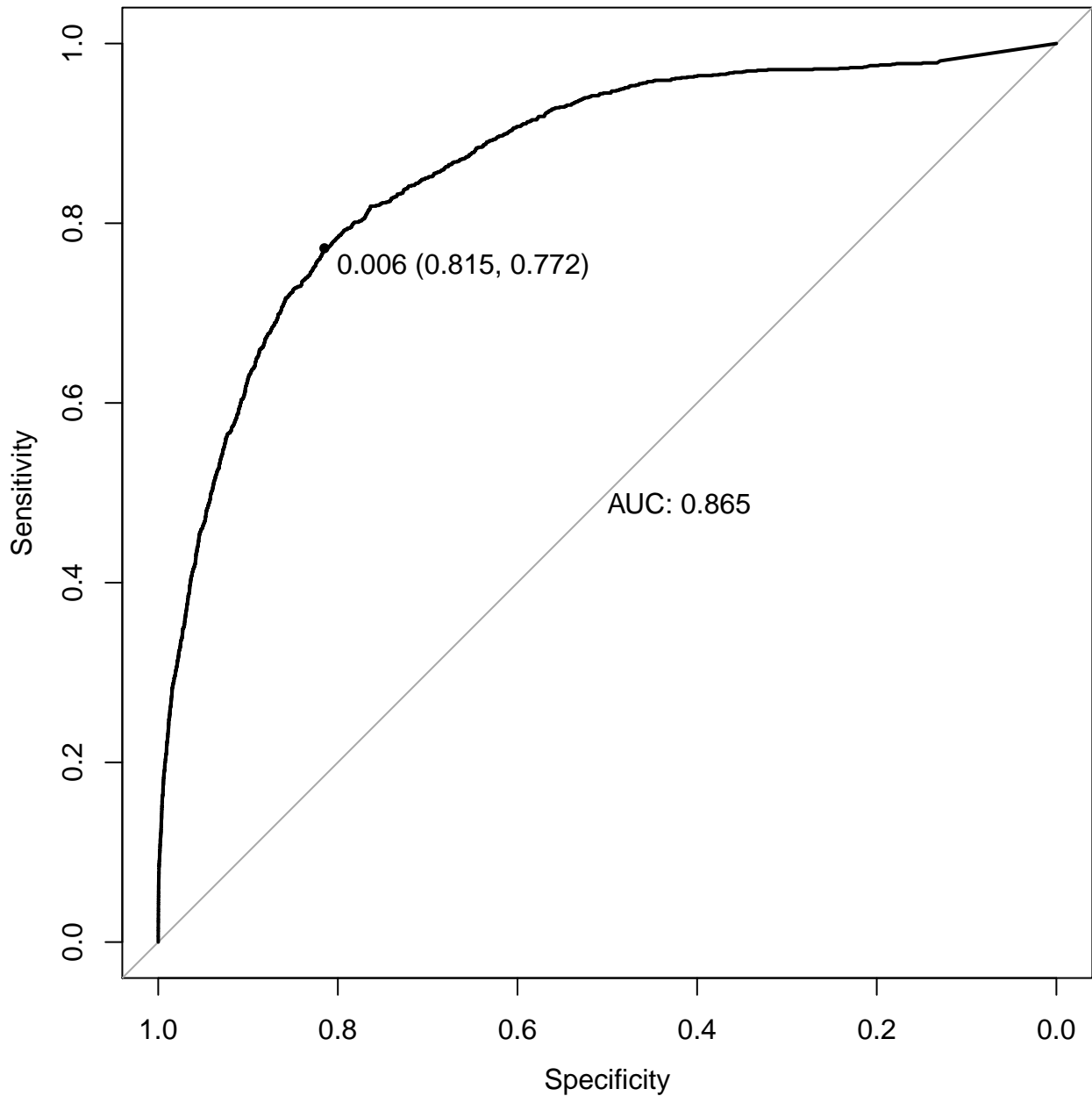
Detection Rate : 0.001262

Detection Prevalence : 0.007512

Balanced Accuracy : 0.579501

'Positive' Class : TRUE

ROC curve for model 13



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1027  -0.0919  -0.0902   3.3217

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.5127573   0.0258341  -213.39  <2e-16 ***
dx_revenue      0.1836691   0.0072724   25.26  <2e-16 ***
dx_session_time  0.0069760   0.0001119   62.36  <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: 30587  on 350797  degrees of freedom
Residual deviance: 26317  on 350795  degrees of freedom
AIC: 26323

Number of Fisher Scoring iterations: 8

```

```
$auc
```

```
Area under the curve: 0.8647
```

```
$relativeCountDifference
```

```
[1] 1.244229
```

```
$optimal_cut_off
```

```
[1] 0.0654198
```


\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173237	1071
TRUE	1399	272

Accuracy : 0.986

95% CI : (0.9854, 0.9865)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.1735

McNemar's Test P-Value : 4.717e-11

Sensitivity : 0.202532

Specificity : 0.991989

Pos Pred Value : 0.162777

Neg Pred Value : 0.993856

Prevalence : 0.007632

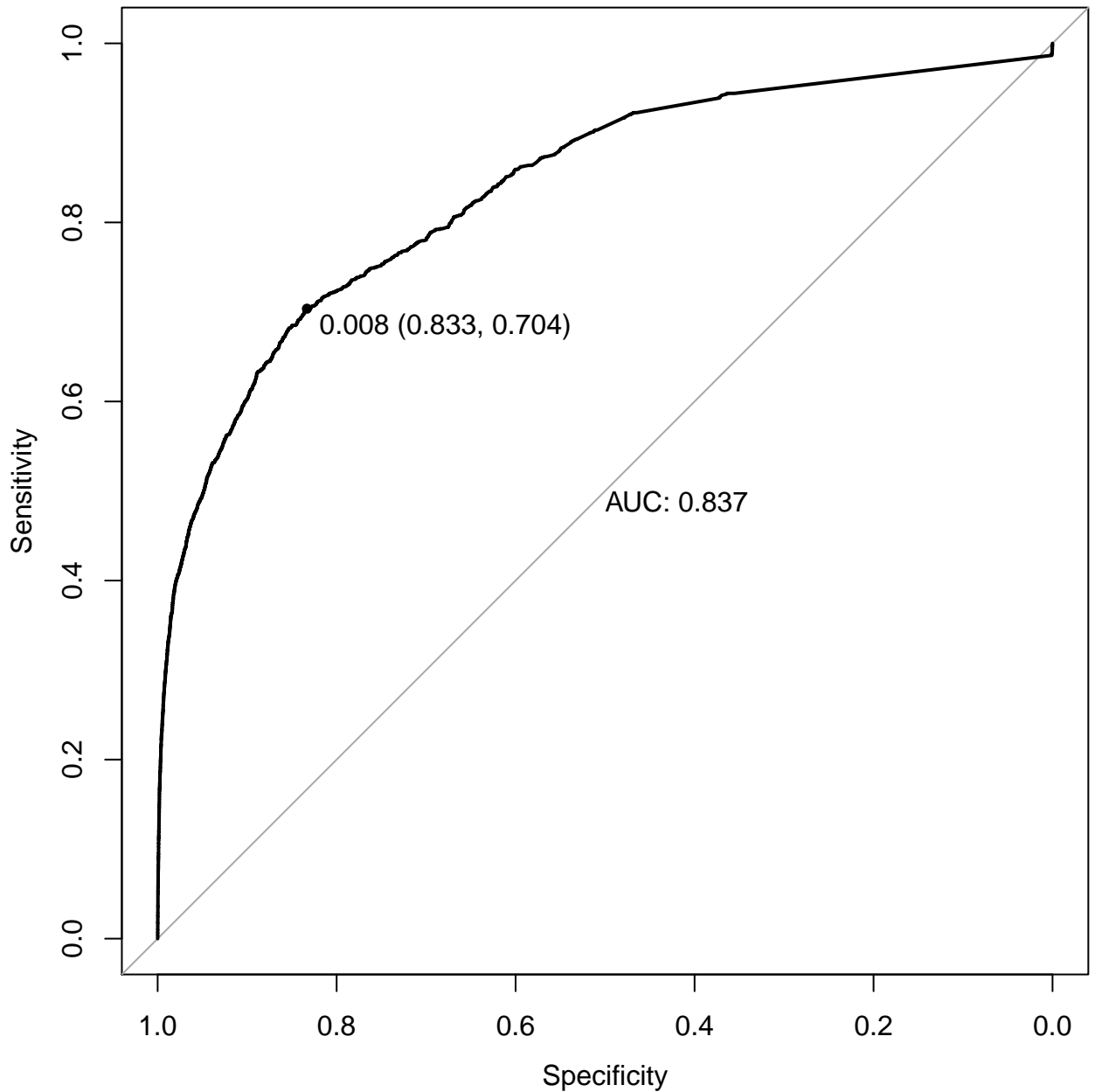
Detection Rate : 0.001546

Detection Prevalence : 0.009495

Balanced Accuracy : 0.597260

'Positive' Class : TRUE

ROC curve for model 14



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1073  -0.0735  -0.0717   5.3017

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.0121059   0.0320116  -187.81  <2e-16 ***
dx_revenue    -0.2992835   0.0127063   -23.55  <2e-16 ***
dx_login_count  0.0492303   0.0024802    19.85  <2e-16 ***
dx_gems_spent  0.0083292   0.0001791    46.51  <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: 30587  on 350797  degrees of freedom
Residual deviance: 24694  on 350794  degrees of freedom
AIC: 24702

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8375
```

```
$relativeCountDifference
```

```
[1] 1.665674
```

```
$optimal_cut_off
```

```
[1] 0.09093946
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	172820	922
TRUE	1816	421

Accuracy : 0.9844
95% CI : (0.9839, 0.985)
No Information Rate : 0.9924
P-Value [Acc > NIR] : 1

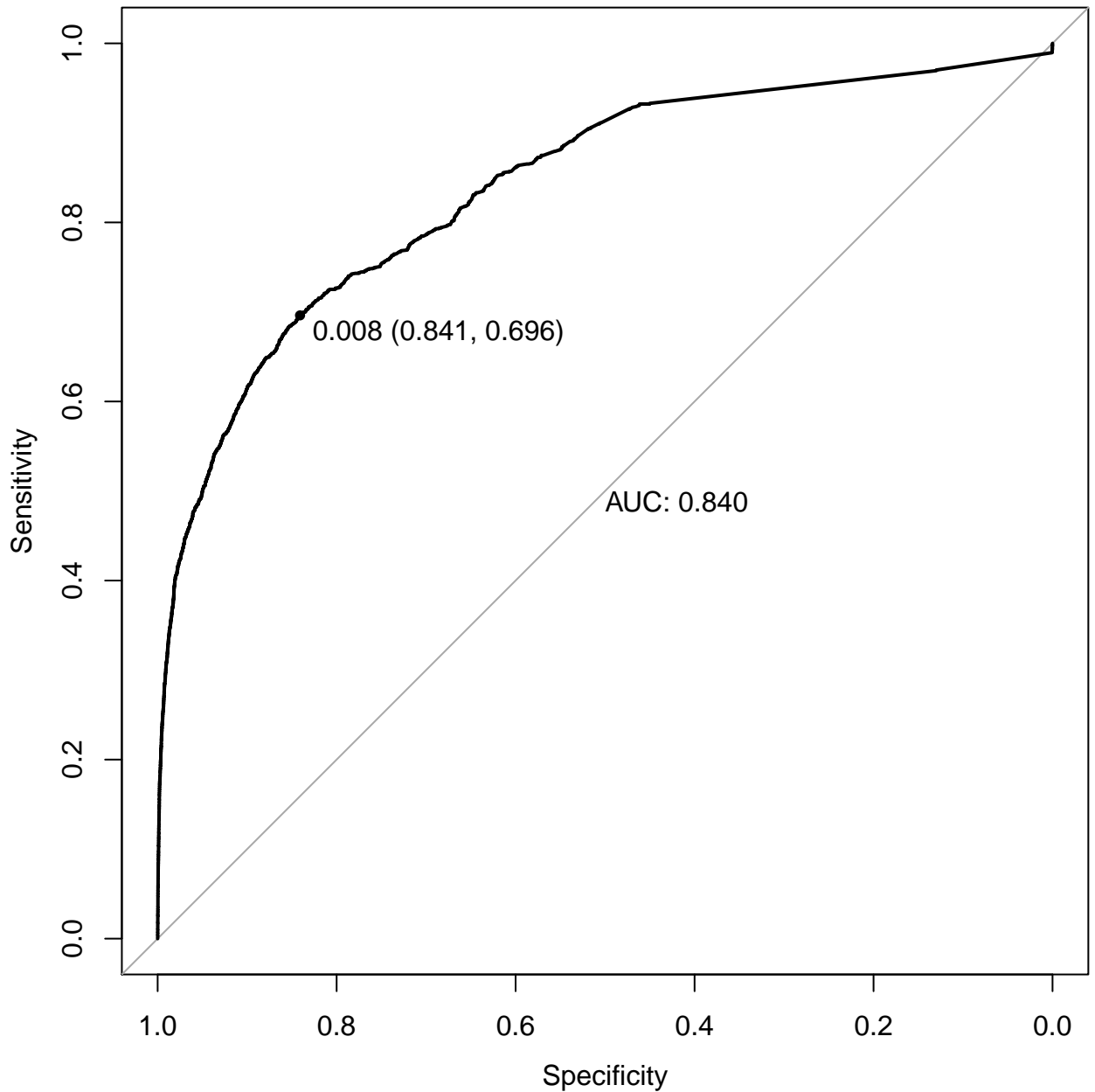
Kappa : 0.2278

McNemar's Test P-Value : <2e-16

Sensitivity : 0.313477
Specificity : 0.989601
Pos Pred Value : 0.188198
Neg Pred Value : 0.994693
Prevalence : 0.007632
Detection Rate : 0.002392
Detection Prevalence : 0.012712
Balanced Accuracy : 0.651539

'Positive' Class : TRUE

ROC curve for model 15



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1066  -0.0720  -0.0720   5.2060

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.0100137  0.0318781 -188.53  <2e-16 ***
dx_revenue    -0.2804122  0.0128507  -21.82  <2e-16 ***
dx_session_count 0.0572301  0.0024817   23.06  <2e-16 ***
dx_gems_spent   0.0079622  0.0001814   43.89  <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: 30587  on 350797  degrees of freedom
Residual deviance: 24590  on 350794  degrees of freedom
AIC: 24598

Number of Fisher Scoring iterations: 9

```



```
$auc
```

```
Area under the curve: 0.8401
```

```
$relativeCountDifference
```

```
[1] 1.597171
```

```
$optimal_cut_off
```

```
[1] 0.09350873
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	172906	928
TRUE	1730	415

Accuracy : 0.9849

95% CI : (0.9843, 0.9855)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.2307

McNemar's Test P-Value : <2e-16

Sensitivity : 0.309010

Specificity : 0.990094

Pos Pred Value : 0.193473

Neg Pred Value : 0.994662

Prevalence : 0.007632

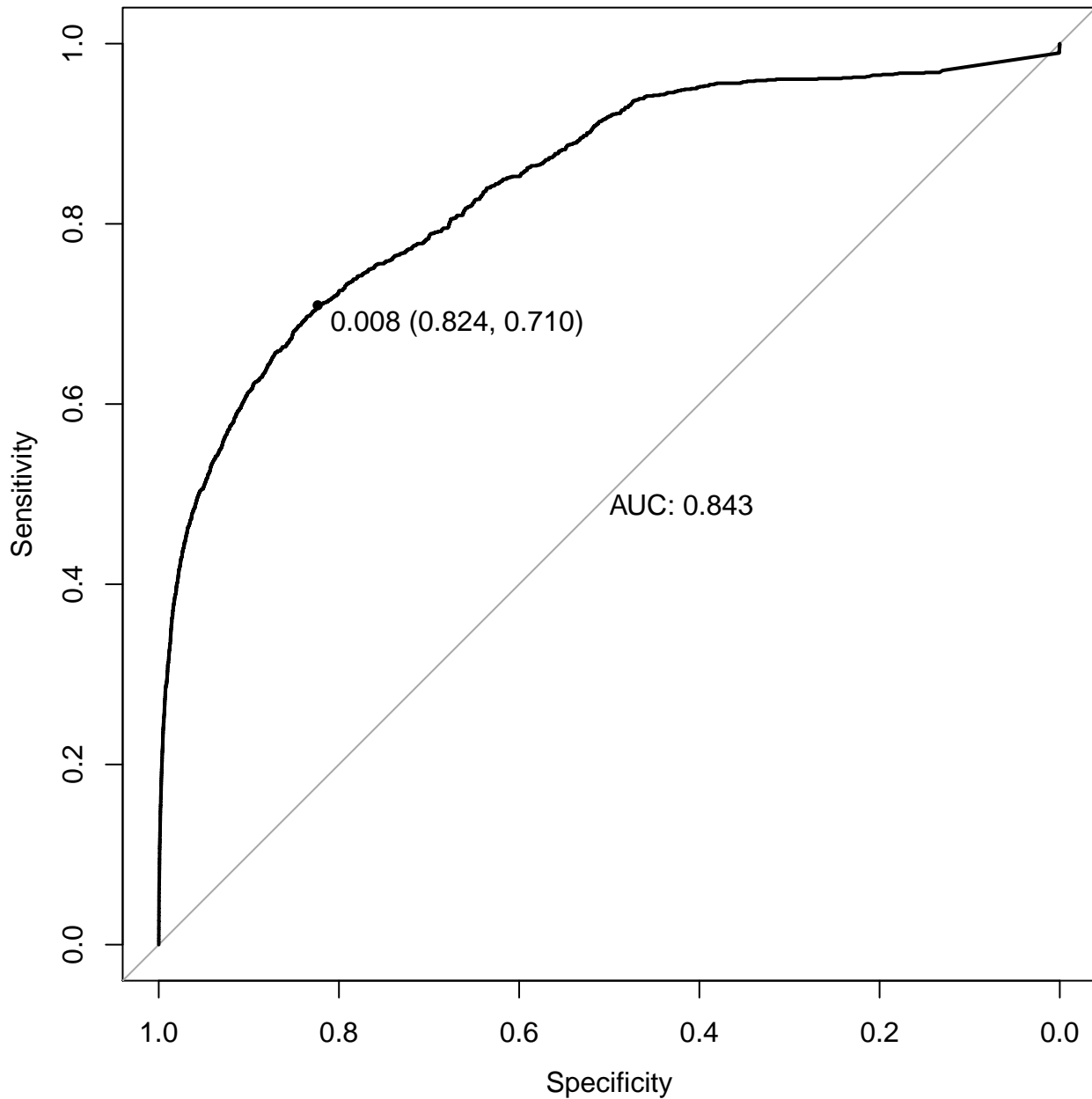
Detection Rate : 0.002358

Detection Prevalence : 0.012189

Balanced Accuracy : 0.649552

'Positive' Class : TRUE

ROC curve for model 16



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1063  -0.0738  -0.0728   5.1660

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.9371828   0.0309719  -191.70  <2e-16 ***
dx_revenue    -0.2731910   0.0133038   -20.54  <2e-16 ***
dx_session_time 0.0034608   0.0001597    21.68  <2e-16 ***
dx_gems_spent  0.0077811   0.0001921    40.52  <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: 30587  on 350797  degrees of freedom
Residual deviance: 24638  on 350794  degrees of freedom
AIC: 24646

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8428
```

```
$relativeCountDifference
```

```
[1] 1.620998
```

```
$optimal_cut_off
```

```
[1] 0.09556242
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	172876	926
TRUE	1760	417

Accuracy : 0.9847

95% CI : (0.9842, 0.9853)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.2297

McNemar's Test P-Value : <2e-16

Sensitivity : 0.310499

Specificity : 0.989922

Pos Pred Value : 0.191548

Neg Pred Value : 0.994672

Prevalence : 0.007632

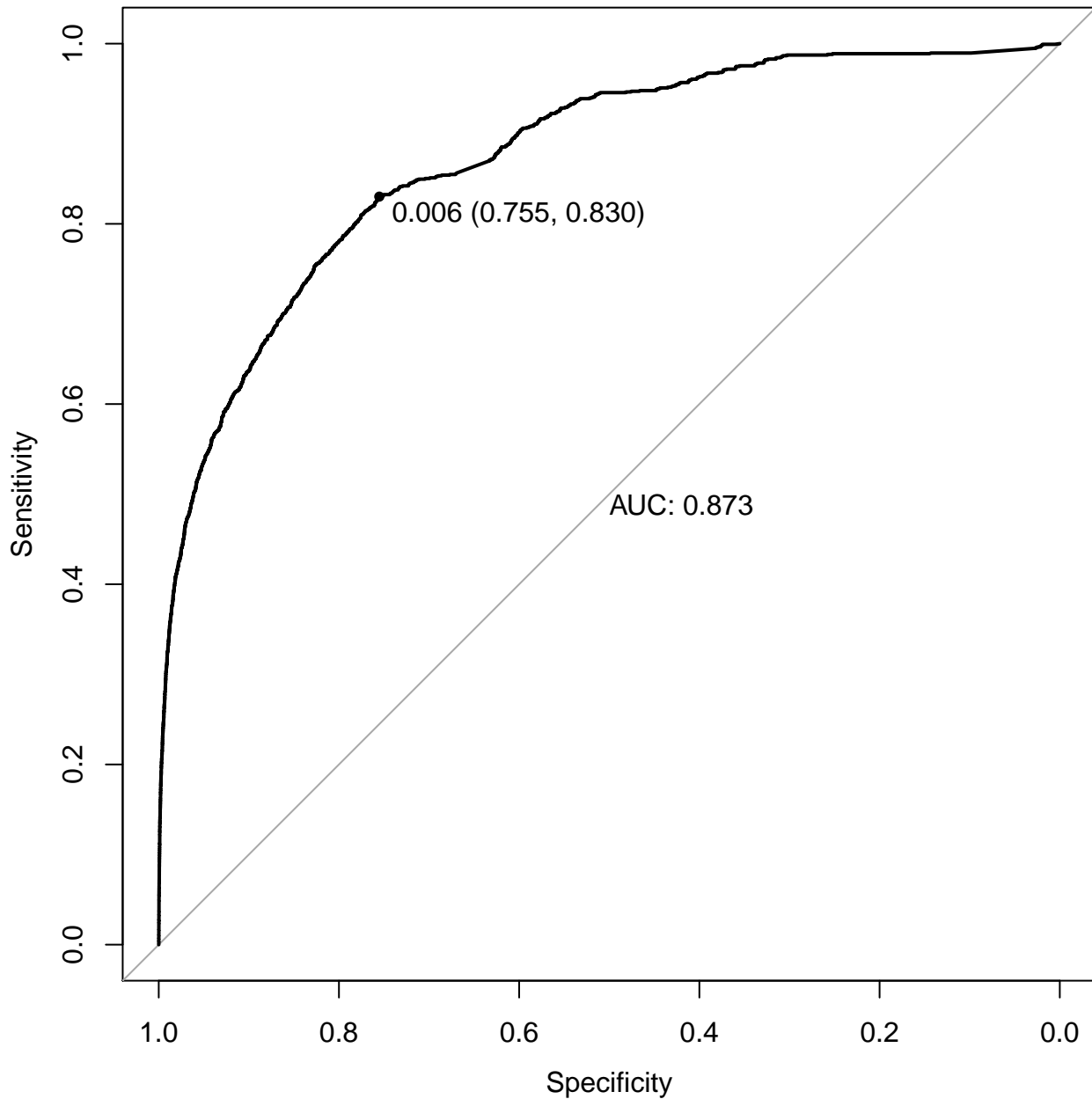
Detection Rate : 0.002370

Detection Prevalence : 0.012371

Balanced Accuracy : 0.650210

'Positive' Class : TRUE

ROC curve for model 17



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.1084  -0.0728  -0.0491   4.9037

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.3485247  0.0524404  -82.92  <2e-16 ***
dx_revenue    -0.2534256  0.0130560  -19.41  <2e-16 ***
dx_session_time 0.0035859  0.0001588   22.58  <2e-16 ***
dx_gems_spent  0.0070868  0.0001898   37.34  <2e-16 ***
tier          -0.7942109  0.0254475  -31.21  <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: 30587  on 350797  degrees of freedom
Residual deviance: 23540  on 350793  degrees of freedom
AIC: 23550

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8732
```

```
$relativeCountDifference
```

```
[1] 1.501862
```

```
$optimal_cut_off
```

```
[1] 0.1051941
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173040	922
TRUE	1596	421

Accuracy : 0.9857

95% CI : (0.9851, 0.9862)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.2437

McNemar's Test P-Value : <2e-16

Sensitivity : 0.313477

Specificity : 0.990861

Pos Pred Value : 0.208726

Neg Pred Value : 0.994700

Prevalence : 0.007632

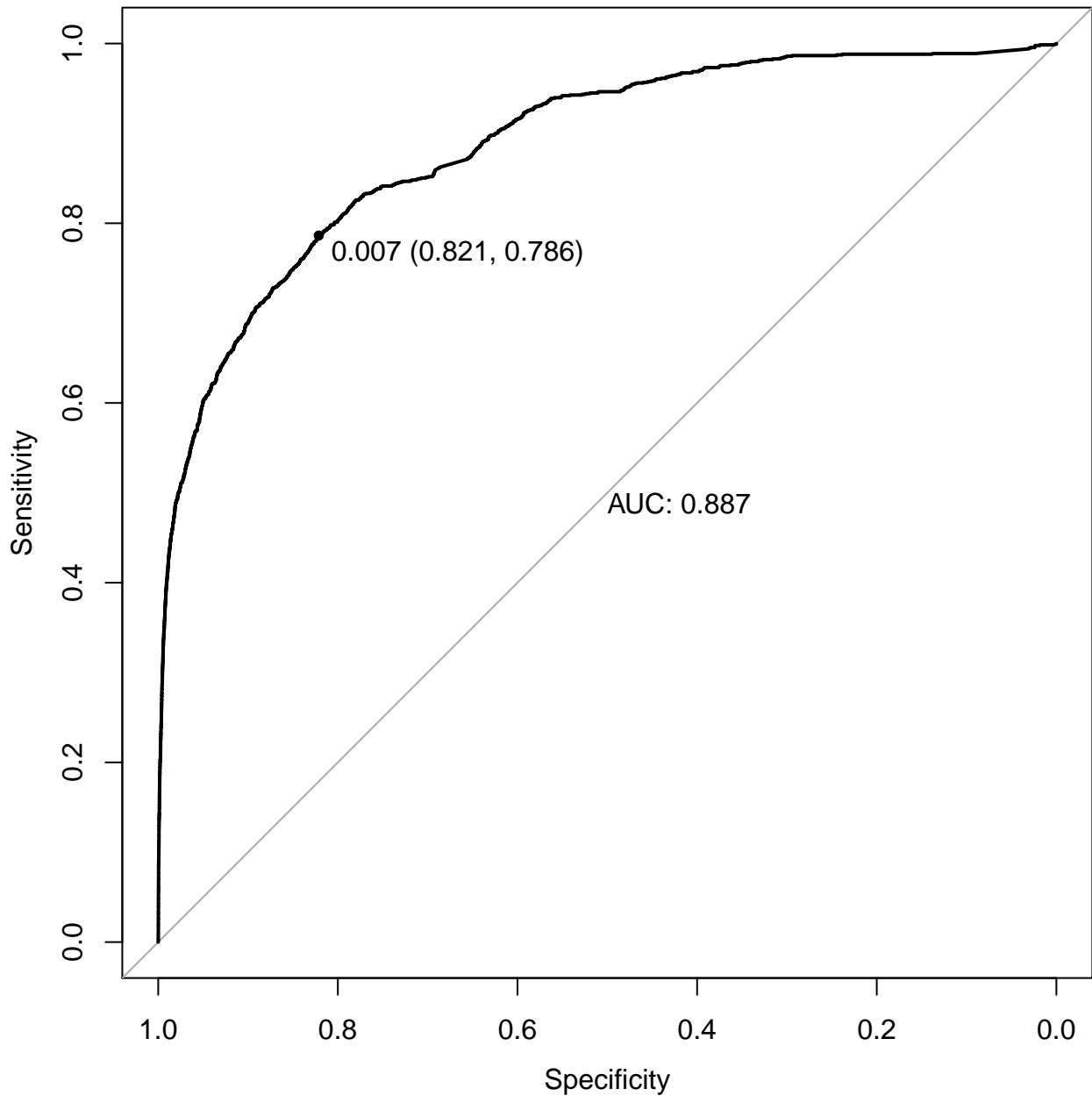
Detection Rate : 0.002392

Detection Prevalence : 0.011462

Balanced Accuracy : 0.652169

'Positive' Class : TRUE

ROC curve for model 18



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0940  -0.1051  -0.0707  -0.0515   4.8749

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.5108642  0.0551144 -81.845 < 2e-16 ***
dx_revenue    -0.2507288  0.0128671 -19.486 < 2e-16 ***
dx_pay_count   2.2223633  0.0548586  40.511 < 2e-16 ***
dx_login_count  0.0063327  0.0070656   0.896  0.37011
dx_session_count 0.0234969  0.0075924   3.095  0.00197 **
dx_session_time  0.0030163  0.0002133  14.143 < 2e-16 ***
dx_gems_count  -0.0007849  0.0018260  -0.430  0.66730
dx_gems_spent   0.0032995  0.0002241  14.723 < 2e-16 ***
tier          -0.7182448  0.0262536 -27.358 < 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: 30587  on 350797  degrees of freedom
Residual deviance: 21794  on 350789  degrees of freedom
AIC: 21812

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8869
```

```
$relativeCountDifference
```

```
[1] 1.216679
```



```
$optimal_cut_off
```

```
[1] 0.1342457
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	173472	873
TRUE	1164	470

Accuracy : 0.9884

95% CI : (0.9879, 0.9889)

No Information Rate : 0.9924

P-Value [Acc > NIR] : 1

Kappa : 0.31

McNemar's Test P-Value : 1.315e-10

Sensitivity : 0.349963

Specificity : 0.993335

Pos Pred Value : 0.287638

Neg Pred Value : 0.994993

Prevalence : 0.007632

Detection Rate : 0.002671

Detection Prevalence : 0.009285

Balanced Accuracy : 0.671649

'Positive' Class : TRUE