Predicting Readmission Probability for Diabetes Inpatients

Executive Summary

In this project, we explored how to help better manage diabetes patients who have been admitted to a hospital. Our goal is to avoid patients being readmitted within 30 days of discharge, which reduces costs for the hospital and improves outcomes for patients. If we could identify important factors relating to the chance of a patient being readmitted within 30 days of discharge, effective intervention could be done to reduce the chance of being readmitted. Also if we could predict one's chance of being readmitted, health professionals can better evaluate who is at risk and take action.

We used a dataset containing over 100,000 different hospital admissions of diabetes patients over a span of 10 years. We used LASSO to select the variables for our logistic regression model. We then were able to create multiple models based off of different sets of coefficients chosen by LASSO with different lambda values and we used anova and AUC values to compare multiple models. The model we selected with the lowest chance of error included 16 different variables that were split between, patient history, characteristics of the patient, lab results, and two medications taken by the patient.

In the end, our model performed fairly well out of a sample, given a train-test-validation split and a higher weight on incorrectly missing patients who will be readmitted in less than 30 days. However, we did still see a good amount of false negatives with the test and validation predictions which we would hope to further refine with future studies. We notably found relatively high positive coefficients associated with previous in-patient visits, discharge type, as well as significant effects associated with primary, secondary, and tertiary diagnosis codes. From that we recommend both further observation and analysis of these features, as well as a reduction in the levels of the diagnosis codes.

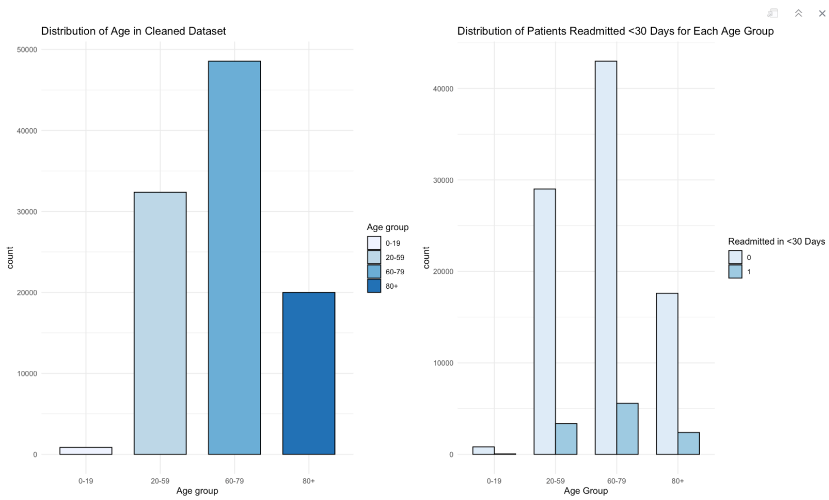
Further limitations included our own lack of background with diabetes and healthcare which caused us to rely more on LASSO for feature selection as we lacked the expertise to rule out variables that make less sense in practice. Additionally, given the large amount of variables used, we do run the risk of overfitting on the training data, although the cross validation performed by LASSO in feature selection does help us avoid that.

Data Summary / EDA

The original data is from the [Center for Clinical and Translational Research] (https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008) at Virginia Commonwealth University. It covers data on diabetes patients across 130 U.S. hospitals from 1999 to 2008. There are over 100,000 unique hospital admissions in this dataset, from ~70,000 unique patients. The data includes demographic elements, such as age, gender, and race, as well as clinical attributes such as tests conducted, emergency/inpatient visits, etc.

For our analysis we will use a cleaned subset of this dataset which excludes variables with lots of missing values and variables with very little variability, as well as applying binning on some categorical variables. The event of interest which we hope to predict with our model is readmitted within < 30 days, so we added a dummy variable indicating whether or not this occurred for a particular patient. .

From the dataset that we are using to create our model we note that there are 11,357 patients that were readmitted in less than 30 days out of the full 101,766 patients that we have data for. We performed some EDA to get a better sense of the makeup of our data, particularly regarding the demographic makeup and the distribution of the response variable (lessThanThirty). More visualizations are present in the Appendix.

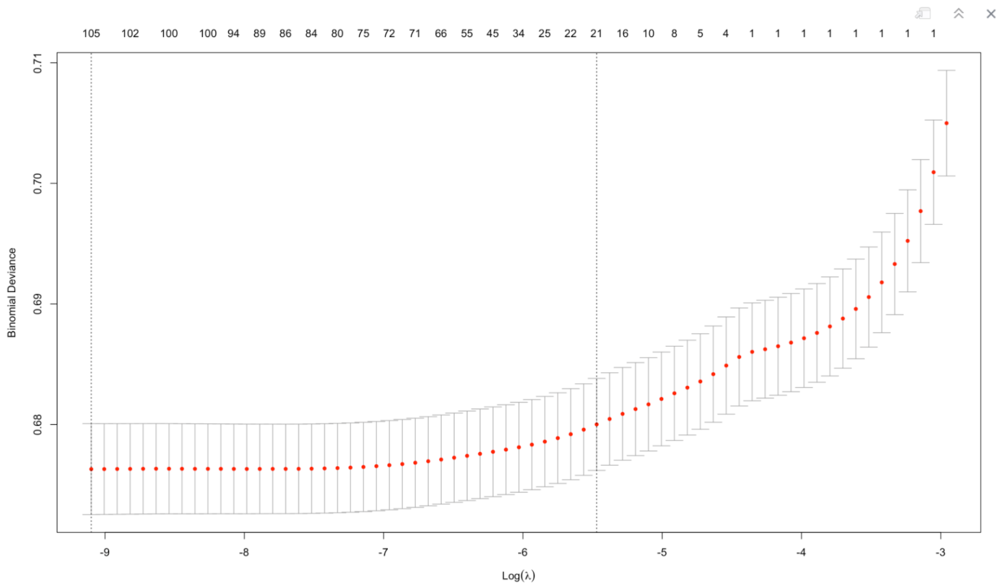


Most notable of the demographic variables is the age group metric, which logically will likely have an effect on an individual's health and hospitalizations. Interestingly, there is a somewhat similar distribution of the response variable with each age group, although it seems to be slightly proportionally higher with older patients.

After looking at the data set, we decided that we needed to modify some of the variables. We started by removing individually unique values such as patient and encounter identifiers as well as the readmitted variable that was no longer needed after creating the lessThanThirty indicator. Our next step was to clean the remaining data. Since we don't have access to additional records for the patients, we decided to remove the patients in which there was missing information, such as those whose race or diag3_mod was coded as "?" or those with a gender value of "Unknown/Invalid". Due to the large size of our data set, removing these data points which amount to only 3,623 rows, should not have a great impact on the analysis and the final model. We also noticed that some variables have an "Other" factor value, which, while it's not easily interpretable, was purposefully placed in the data and doesn't constitute an error or missing value.

Analysis and Model Selection

With that finished, we were able to begin the variable selection process for our model. We performed logistic regression using cv.glmnet on the 28 variables that we had.



The plot above shows us the ideal value of lambda to use for the model. We ran two different values of lambda to get the coefficients for different models. The coefficients that LASSO selected for the model using lambda.1se are as follows:

```
[1] "(Intercept)"
[2] "time_in_hospital"
[3] "num_medications"
[4] "number_emergency"
[5] "number_inpatient"
[6] "number_diagnoses"
[7] "insulinNo"
[8] "diabetesMedYes"
[9] "disch_disp_modifiedDischarged to
home with Home Health Service"
[10]"disch_disp_modifiedDischarged/Tran
sferred to SNF"
[11] "disch_disp_modifiedOther"
[12] "age_mod20-59"
[13] "diag1_mod428"
[14] "diag1_mod434"
[15] "diag1_mod486"
[16] "diag1_mod786"
[17] "diag1_mod820"
[18] "diag2_mod401"
[19] "diag3_mod250.6"
[20] "diag3_mod401"
[21] "diag3_mod403"
[22] "diag3_mod585"
```

Lambda.1se is the value of lambda that gives the most regularized model such that the cross-validated error is within one standard error of the minimum. We have 12 different overall variables that were selected with some of them having multiple levels in this model, there are a total of 21 coefficients and the intercept.

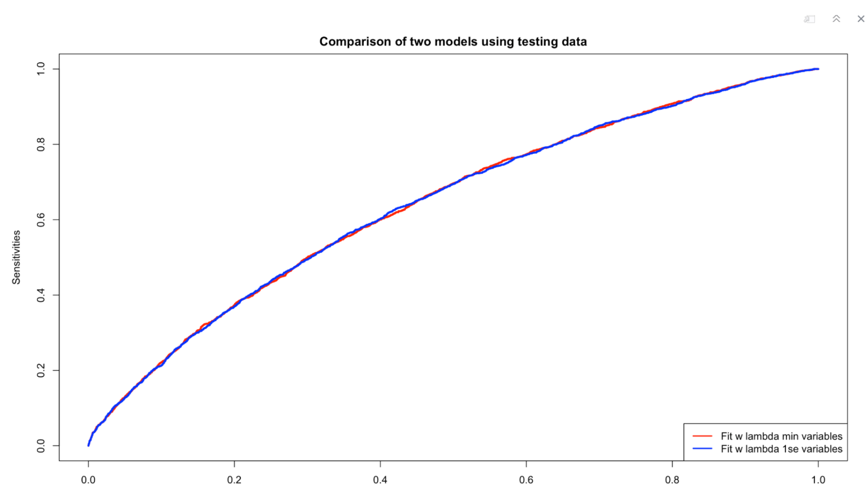
Appendix Output I gives the 106 coefficients (including the intercept) that LASSO selected using lamda.min. Which is the value of lambda that gives minimum mean cross-validated error. We can note that many more variables were selected in this model with additional levels for some of the factors that were included in the 1se model. We can see that 28 variables were selected for this model, which is every single variable that we are using in this model, however not all of the levels were included for each of the factor variables.

Next, we fit initial logistic regression models using the two different groups of variables that LASSO selected. For both models, since we have factor variables in which not all of the levels were selected, we decided to force all of the levels in and use Anova to determine if the categorical variables were in fact significant.

For lambda min, we found that at a significance level of 0.05, 16 of the 28 variables were significant (*Appendix Output II*). We then created a model with just those 16 variables to compare to the initial larger model using anova, in order to determine if the dropped variables create a significant difference. In running anova on the reduced model and the full model, we saw that there was no significant difference between the two different models (*Appendix Output III*). Since there isn't a significant difference, we will select the model with fewer variables to continue on with our analysis, and for robustness, using Anova (*Appendix Output V*) we were able to determine that all the 16 variables were significant in a logistic regression model only using them.

Moving onto the lambda se model, we again used Anova on a logistic regression model using the selected variables to determine which were significant. We found that all of the variables except for time in hospital are significant in the model (*Appendix Output VI*). We then decided to compare this model to the prior one that we tested for lambda min using anova (*Appendix Output VII*). We get that for these two models, the p-value is very low (2.48e-07) meaning there is a significant difference between the two models. The key variables that are included in the model using lambda min are: num_procedures, A1Cresult, metformin, glipizide, and adm_src_mod.

Given that the models were significantly different, to decide between the two different models, we plotted their ROC curves and compared their AUC values. Before we were able to compute these curves, we first split the dataset into training, testing and validation sets using a 60-20-20 split. We then restrained new logistic regression models using the parameters as selected previously on the new training set. Using the predictions of these models on the testing set, we found these ROC curves and AUC values:



We can see that the AUC values and ROC curves are quite similar, but given that the slightly larger model using the variables found using lambda min has a higher AUC value, we decided to use that. We also consulted a nurse who endorsed the fact that A1c results are an important indicator for a diabetes patient and should be included. The summary statistics of our final model trained on the training data are shown in Output VIII of the Appendix. It's important to note that although some variables are not shown to be significant with the training set, we still kept them in given that LASSO performs k-folds validation in order to pick variables. The train-test split we performed was utilized more to evaluate the performance of our model and determine a threshold for the probability predictions returned by our logistic regression model. The final parameters used for our model were:

- num_procedures
- number_emergency
- number_diagnoses
- metformin
- insulin
- disch_disp_modified
- Age_mod
- diag2_mod
- num_medications
- number_inpatient
- A1Cresult
- glipizide
- diabetesMed
- adm_src_mod
- diag1_mod
- diag3_mod

In determining the threshold to use for prediction, we first graphed the false positive rate for the test set over different thresholds to get a sense of how performance changed with different thresholds (*Figure III*). The graph indicated a threshold of around 0.2 would offer a good tradeoff between actually predicting quick readmissions but not so many that there's a large false positive rate.

In determining a more exact loss function and threshold, we utilized a weighted classification error, given that it is more harmful to fail to predict that a patient will be readmitted in less than 30 days rather than falsely predict that they will. A false negative could dissuade medical professionals from checking in on a patient who is at risk and may not be able to get to the hospital when they need to within a few weeks later. Therefore, we assigned an estimated weight where false negatives, mislabeling a readmission ($a_{1,0} = L(Y = 1, \widehat{Y} = 0)$) were worth twice as much as false positives, or mislabeling a non-readmission ($a_{0,1} = L(Y = 0, \widehat{Y} = 1)$).

We can then calculate a threshold of:

$$\widehat{P}(final\ model\ variables) > \frac{0.5}{1 + 0.5} \approx 0.167$$

Using this threshold, we get a weighted misclassification error (MSE) of 0.229 on the test set and 0.231 on the validation set which isn't too bad of a result.

To understand these results a bit better, we created confusion matrices for the test and validation sets:

Test set confusion matrix:

Actual

		0	1
Predicted	0	17,218	2,178
	1	147	85

Validation set confusion matrix:

		Actual	
		0	1
Predicted	0	17,229	2,218
	1	102	81

We can see that the quantity of false negatives in both matrices likely contributed to higher MSE values.

Recommendations

Our primary recommendation is focused on the number_inpatient and disch_disp_modified variables given their statistical significance, interpretability, and large positive coefficients. We recommend that hospital staff closely watch diabetes patients with higher counts of inpatient visits each year, as it doesn't seem to be isolated instances and positively affects their chance of being shortly readmitted during the next year. Additionally, given that every level of the variable indicating where the patient was discharged was significant, we recommend that they delve more into the "Other" category given its high coefficient and look into those who are simply discharged home (the baseline variable for these high coefficients).to determine if they are in fact in less risk of being readmitted or don't have anyone to take care of them and worry enough to bring them to the hospital.

To improve this study, it would be interesting to see how our model performs in predicting diabetes patient readmissions within 30 days in the years after 2008 and whether the chosen features remain significant. The number of hospitals sampled could also be expanded to analyze these results as well. We could also find ways to incorporate other features such as some regarding the patient's living conditions or whether they have been readmitted to the hospital in the short time frame in the past.

It would also help to reduce the number of levels associated with the primary, secondary, and tertiary diagnoses by binning them into groups, given that diag2_mod alone has 25 levels. With the advice of a medical professional with more expertise on diabetes and these associated diagnoses, this could improve our model. A medical professional could also provide more insight on the significance of the two medications that were found to be significant in our model and give better insight into what implications this may have on patient care.

Appendix 1: Figures

Figure I

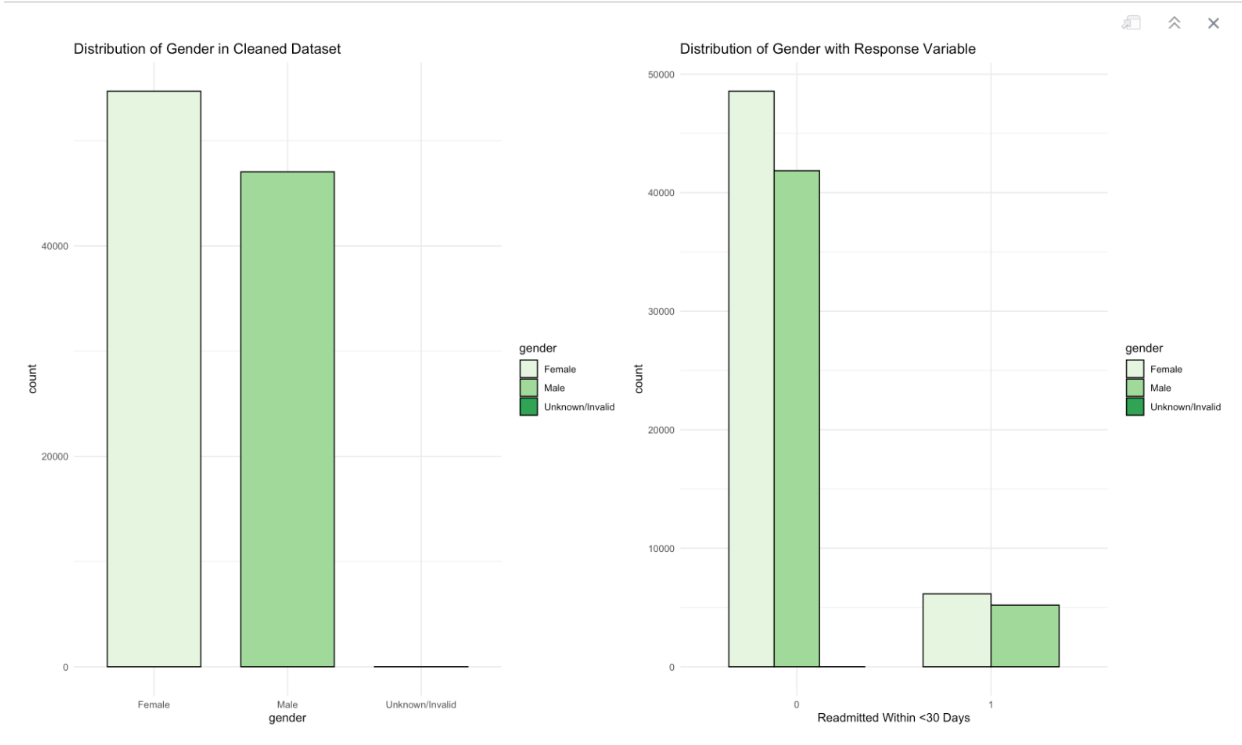


Figure II

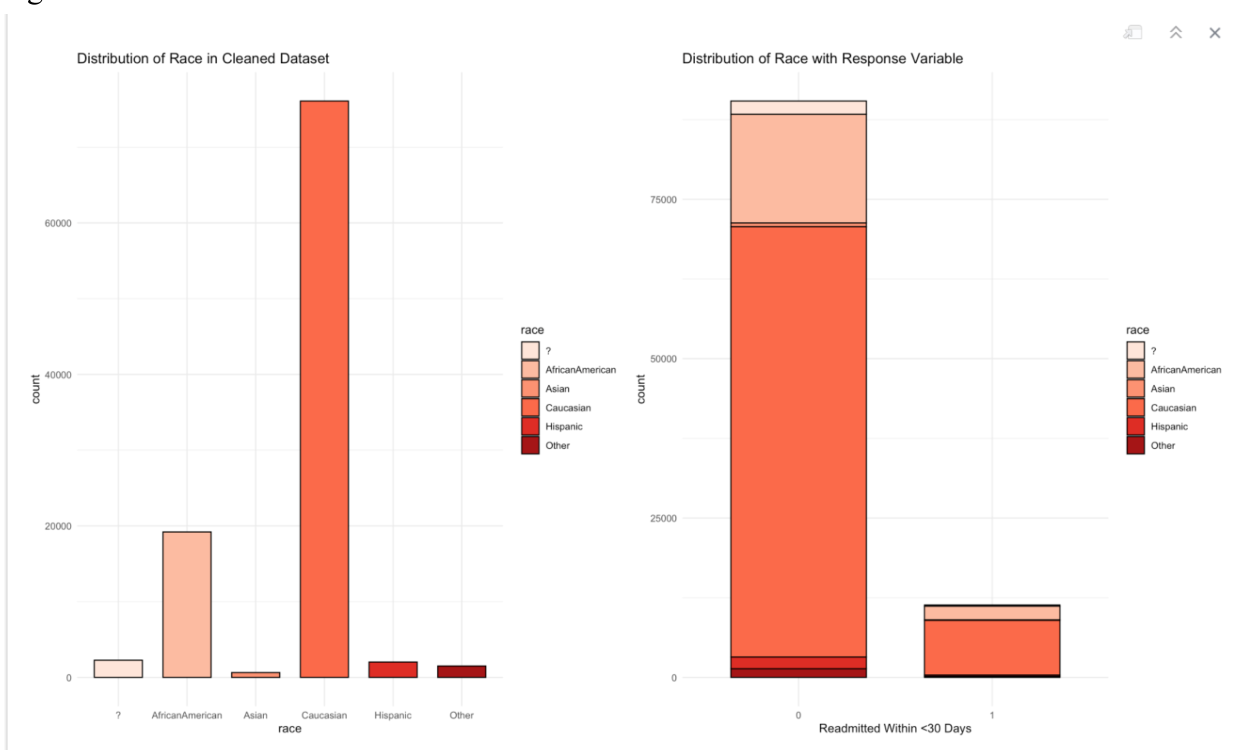
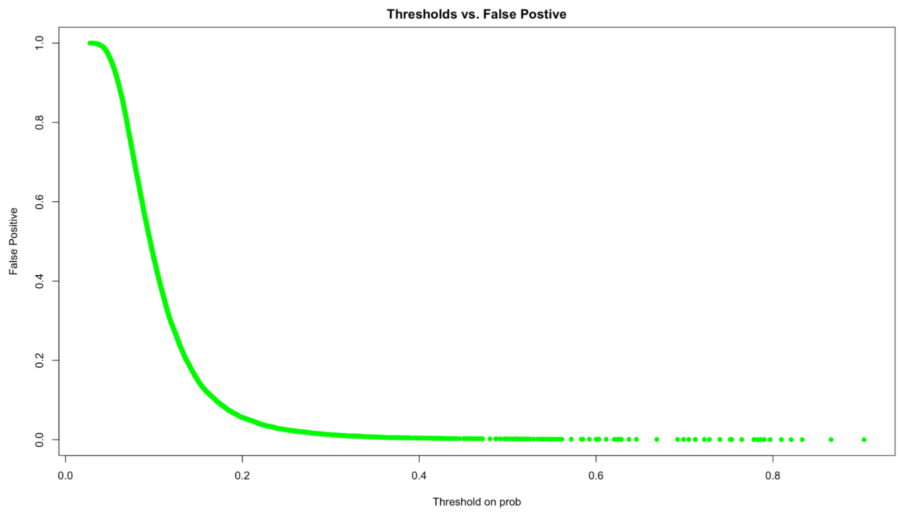


Figure III


Thresholds vs. False Postive
False Positive
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
Threshold on prob

Appendix 2: Outputs

Output I

```
 [1] "(Intercept)"
 [2] "raceCaucasian"
 [3] "raceOther"
 [4] "genderMale"
 [5] "time_in_hospital"
 [6]"num_lab_procedures"
 [7] "num_procedures"
 [8] "num_medications"
 [9] "number_outpatient"
 [10] "number_emergency"
 [11] "number_inpatient"
 [12] "number_diagnoses"
 [13]"max_glu_serumNone"
 [14] "A1CresultNone"
 [15] "A1CresultNorm"
 [16] "metforminSteady"
 [17] "metforminUp"
 [18]"glimepirideSteady"
 [19] "glipizideNo"
 [20] "glipizideSteady"
 [21] "glyburideNo"
 [22] "glyburideSteady"
 [23] "glyburideUp"
[24]"pioglitazoneSteady"
 [25] "pioglitazoneUp"
 [26] "rosiglitazoneNo"
 [27] "rosiglitazoneUp"
 [28] "insulinNo"
 [29] "insulinSteady"
 [30] "insulinUp"
 [31] "changeNo"
 [32] "diabetesMedYes"
[33]"disch_disp_modified
Discharged to home with
Home Health Service"
[34]"disch_disp_modified
Discharged/Transferred
to SNF"
[35]"disch_disp_modified
Other"
 [36] "adm_src_modOther"
[37]"adm_src_modPhysicia
n Referral"
[38]"adm_src_modTransfer
from Home Health"
[39]"adm_typ_modEmergenc
y"
 [40]"adm_typ_modUrgent"
 [41] "age_mod20-59"
 [42] "age_mod60-79"
 [43] "age_mod80+"
 [44] "diag1_mod250.8"
 [45] "diag1_mod276"
 [46] "diag1_mod38"
 [47] "diag1_mod410"
 [48] "diag1_mod414"
 [49] "diag1_mod427"
 [50] "diag1_mod434"
 [51] "diag1_mod435"
 [52] "diag1_mod486"
 [53] "diag1_mod491"
 [54] "diag1_mod493"
 [55] "diag1_mod518"
 [56] "diag1_mod577"
 [57] "diag1_mod584"
 [58] "diag1_mod599"
 [59] "diag1_mod682"
 [60] "diag1_mod715"
 [61] "diag1_mod780"
 [62] "diag1_mod786"
 [63] "diag1_mod820"
 [64] "diag1_mod996"
 [65] "diag1_modOther"
 [66] "diag2_mod250.01"
 [67] "diag2_mod250.02"
 [68] "diag2_mod276"
 [69] "diag2_mod285"
 [70] "diag2_mod401"
 [71] "diag2_mod403"
 [72] "diag2_mod411"
 [73] "diag2_mod413"
 [74] "diag2_mod414"
 [75] "diag2_mod424"
 [76] "diag2_mod425"
 [77] "diag2_mod427"
 [78] "diag2_mod428"
 [79] "diag2_mod486"
 [80] "diag2_mod491"
 [81] "diag2_mod518"
 [82] "diag2_mod584"
 [83] "diag2_mod585"
 [84] "diag2_mod599"
 [85] "diag2_mod682"
 [86] "diag2_mod707"
 [87] "diag2_mod780"
 [88] "diag2_modOther"
 [89] "diag3_mod250.02"
 [90] "diag3_mod250.6"
 [91] "diag3_mod272"
 [92] "diag3_mod276"
 [93] "diag3_mod285"
 [94] "diag3_mod401"
 [95] "diag3_mod403"
 [96] "diag3_mod414"
 [97] "diag3_mod425"
 [98] "diag3_mod427"
 [99] "diag3_mod428"
[100] "diag3_mod496"
[101] "diag3_mod585"
[102] "diag3_mod599"
[103] "diag3_mod707"
[104] "diag3_mod780"
[105] "diag3_modOther"
[106] "diag3_modV45"
```

Output II

```
Analysis of Deviance Table (Type II tests)

Response: lessThanThirty
                    LR Chisq Df Pr(>Chisq)
race                    1.79  4   0.774958
time_in_hospital        1.18  1   0.277538
num_lab_procedures      0.54  1   0.461294
```

```
num_procedures          6.57  1   0.010379 *
num_medications         8.98  1   0.002730 **
number_outpatient       0.66  1   0.416627
number_emergency       19.65  1  9.315e-06 ***
number_inpatient     1277.64  1  < 2.2e-16 ***
number_diagnoses        5.90  1   0.015098 *
max_glu_serum           6.79  3   0.078797 .
A1Cresult              11.02  3   0.011615 *
metformin              12.40  3   0.006142 **
glimepiride             5.24  3   0.155008
glipizide               8.89  3   0.030795 *
glyburide               2.04  3   0.564614
pioglitazone            2.20  3   0.531762
rosiglitazone           3.77  3   0.286872
insulin                11.85  3   0.007920 **
change                  0.38  1   0.538156
diabetesMed            31.92  1  1.604e-08 ***
disch_disp_modified   222.24  3  < 2.2e-16 ***
adm_src_mod            15.47  3   0.001455 **
adm_typ_mod             2.28  3   0.517213
age_mod                30.75  3  9.585e-07 ***
diag1_mod             198.65 23  < 2.2e-16 ***
diag2_mod              70.00 24  2.185e-06 ***
diag3_mod              92.58 19  1.152e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output III

```
Analysis of Deviance Table

Model 1: lessThanThirty ~ num_procedures + num_medications + number_emergency +
    number_inpatient + number_diagnoses + A1Cresult + metformin +
    glipizide + insulin + diabetesMed + disch_disp_modified +
    adm_src_mod + age_mod + diag1_mod + diag2_mod + diag3_mod
Model 2: lessThanThirty ~ race + time_in_hospital + num_lab_procedures +
    num_procedures + num_medications + number_outpatient + number_emergency +
    number_inpatient + number_diagnoses + max_glu_serum + A1Cresult +
    metformin + glimepiride + glipizide + glyburide + pioglitazone +
    rosiglitazone + insulin + change + diabetesMed + disch_disp_modified +
    adm_src_mod + adm_typ_mod + age_mod + diag1_mod + diag2_mod +
    diag3_mod
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     98049      66161
2     98023      66132 26   29.003   0.3109
```

Output IV

```
Analysis of Deviance Table (Type II tests)

Response: lessThanThirty
                    LR Chisq Df Pr(>Chisq)
num_procedures          6.95  1  0.0083891 **
num_medications        13.71  1  0.0002129 ***
number_emergency       18.63  1  1.586e-05 ***
```

```
number_inpatient     1301.74  1  < 2.2e-16 ***
number_diagnoses        5.28  1  0.0215434 *
A1Cresult              10.59  3  0.0141615 *
metformin              16.26  3  0.0010043 **
glipizide              10.39  3  0.0155283 *
insulin                14.01  3  0.0028908 **
diabetesMed            27.33  1  1.714e-07 ***
disch_disp_modified   243.60  3  < 2.2e-16 ***
adm_src_mod            13.41  3  0.0038242 **
age_mod                30.88  3  9.014e-07 ***
diag1_mod             199.66 23  < 2.2e-16 ***
diag2_mod              70.91 24  1.592e-06 ***
diag3_mod              95.24 19  3.860e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output V

```
Analysis of Deviance Table (Type II tests)

Response: lessThanThirty
                    LR Chisq Df Pr(>Chisq)
num_procedures          6.95  1  0.0083891 **
num_medications        13.71  1  0.0002129 ***
number_emergency       18.63  1  1.586e-05 ***
number_inpatient     1301.74  1  < 2.2e-16 ***
number_diagnoses        5.28  1  0.0215434 *
A1Cresult              10.59  3  0.0141615 *
metformin              16.26  3  0.0010043 **
glipizide              10.39  3  0.0155283 *
insulin                14.01  3  0.0028908 **
diabetesMed            27.33  1  1.714e-07 ***
disch_disp_modified   243.60  3  < 2.2e-16 ***
adm_src_mod            13.41  3  0.0038242 **
age_mod                30.88  3  9.014e-07 ***
diag1_mod             199.66 23  < 2.2e-16 ***
diag2_mod              70.91 24  1.592e-06 ***
diag3_mod              95.24 19  3.860e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output VI

```
Analysis of Deviance Table (Type II tests)

Response: lessThanThirty
                    LR Chisq Df Pr(>Chisq)
time_in_hospital        0.89  1   0.346657
num_medications         4.27  1   0.038709 *
number_emergency       18.72  1  1.517e-05 ***
number_inpatient     1366.26  1  < 2.2e-16 ***
number_diagnoses        6.24  1   0.012503 *
insulin                15.88  3   0.001199 **
diabetesMed            26.29  1  2.932e-07 ***
disch_disp_modified   232.60  3  < 2.2e-16 ***
```

```
age_mod                34.83  3  1.326e-07 ***
diag1_mod             198.60 23  < 2.2e-16 ***
diag2_mod              71.66 24  1.224e-06 ***
diag3_mod              96.07 19  2.742e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output VII

```
Analysis of Deviance Table

Model 1: lessThanThirty ~ num_procedures + num_medications + number_emergency +
    number_inpatient + number_diagnoses + A1Cresult + metformin +
    glipizide + insulin + diabetesMed + disch_disp_modified +
    adm_src_mod + age_mod + diag1_mod + diag2_mod + diag3_mod
Model 2: lessThanThirty ~ num_medications + number_emergency + number_inpatient +
    number_diagnoses + insulin + diabetesMed + disch_disp_modified +
    age_mod + diag1_mod + diag2_mod + diag3_mod
  Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
1    101671      67999
2    101684      68055 -13  -56.204 2.481e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output VIII

```
Call:
glm(formula = lessThanThirty ~ num_procedures + num_medications +
    number_emergency + number_inpatient + number_diagnoses +
    A1Cresult + metformin + glipizide + insulin + diabetesMed +
    disch_disp_modified + adm_src_mod + age_mod + diag1_mod +
    diag2_mod + diag3_mod, family = binomial, data = data.train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1550  -0.5090  -0.4323  -0.3676   2.6833

Coefficients:
                                                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                 -2.477876   0.374857  -6.610 3.84e-11 ***
num_procedures                                              -0.013408   0.010104  -1.327 0.184524
num_medications                                              0.002291   0.002003   1.143 0.252859
number_emergency                                             0.037529   0.011142   3.368 0.000756 ***
number_inpatient                                             0.248638   0.008575  28.995  < 2e-16 ***
number_diagnoses                                             0.023275   0.009141   2.546 0.010887 *
A1Cresult>8                                                 -0.089433   0.087070  -1.027 0.304356
A1CresultNone                                                0.052903   0.071767   0.737 0.461025
A1CresultNorm                                               -0.014172   0.094277  -0.150 0.880510
metforminNo                                                 -0.375211   0.157862  -2.377 0.017462 *
metforminSteady                                             -0.444456   0.159834  -2.781 0.005424 **
metforminUp                                                 -0.646324   0.215968  -2.993 0.002765 **
glipizideNo                                                 -0.192803   0.165434  -1.165 0.243843
glipizideSteady                                             -0.124586   0.168915  -0.738 0.460779
glipizideUp                                                  0.149267   0.213957   0.698 0.485398
insulinNo                                                   -0.118159   0.048623  -2.430 0.015094 *
insulinSteady                                               -0.108216   0.043636  -2.480 0.013139 *
insulinUp                                                   -0.066530   0.051946  -1.281 0.200280
diabetesMedYes                                               0.178377   0.045024   3.962 7.44e-05 ***
disch_disp_modifiedDischarged to home with Home Health Service  0.205035   0.041488   4.942 7.73e-07 ***
disch_disp_modifiedDischarged/Transferred to SNF             0.368387   0.041436   8.891  < 2e-16 ***
disch_disp_modifiedOther                                     0.423469   0.038141  11.103  < 2e-16 ***
adm_src_modOther                                            -0.061472   0.054624  -1.125 0.260439
adm_src_modPhysician Referral                                0.002029   0.032954   0.062 0.950909
```

```
adm_src_modTransfer from Home Health                          -0.130320   0.056445  -2.309 0.020956 *
age_mod20-59                                                   0.513318   0.248134   2.069 0.038573 *
age_mod60-79                                                   0.632580   0.248490   2.546 0.010906 *
age_mod80+                                                     0.603732   0.249887   2.416 0.015691 *
diag1_mod250.8                                                -0.620116   0.145538  -4.261 2.04e-05 ***
diag1_mod276                                                  -0.374606   0.136753  -2.739 0.006157 **
diag1_mod38                                                   -0.608782   0.145144  -4.194 2.74e-05 ***
diag1_mod410                                                  -0.511383   0.129616  -3.945 7.97e-05 ***
diag1_mod414                                                  -0.499037   0.128511  -3.883 0.000103 ***
diag1_mod427                                                  -0.667810   0.136509  -4.892 9.98e-07 ***
diag1_mod428                                                  -0.332486   0.113200  -2.937 0.003312 **
diag1_mod434                                                  -0.029995   0.130059  -0.231 0.817604
diag1_mod435                                                  -0.807152   0.192001  -4.204 2.62e-05 ***
diag1_mod486                                                  -0.795868   0.129930  -6.125 9.05e-10 ***
diag1_mod491                                                  -0.321712   0.131417  -2.448 0.014364 *
diag1_mod493                                                  -0.608540   0.176060  -3.456 0.000547 ***
diag1_mod518                                                  -1.014853   0.177763  -5.709 1.14e-08 ***
diag1_mod577                                                  -0.243328   0.157297  -1.547 0.121879
diag1_mod584                                                  -0.408858   0.141657  -2.886 0.003899 **
diag1_mod599                                                  -0.587092   0.146189  -4.016 5.92e-05 ***
diag1_mod682                                                  -0.743435   0.144271  -5.153 2.56e-07 ***
diag1_mod715                                                  -0.354557   0.142442  -2.489 0.012806 *
diag1_mod780                                                  -0.712314   0.145904  -4.882 1.05e-06 ***
diag1_mod786                                                  -0.744429   0.131100  -5.678 1.36e-08 ***
diag1_mod820                                                  -0.186893   0.152371  -1.227 0.219987
diag1_mod996                                                  -0.386072   0.133733  -2.887 0.003891 **
diag1_modOther                                                -0.428804   0.102501  -4.183 2.87e-05 ***
diag2_mod250.01                                                0.370670   0.125191   2.961 0.003068 **
diag2_mod250.02                                                0.201027   0.117708   1.708 0.087664 .
diag2_mod276                                                   0.191852   0.089248   2.150 0.031583 *
diag2_mod285                                                  -0.114908   0.136883  -0.839 0.401210
diag2_mod401                                                  -0.054104   0.112000  -0.483 0.629045
diag2_mod403                                                   0.188815   0.106161   1.779 0.075310 .
diag2_mod411                                                   0.067745   0.130558   0.519 0.603840
diag2_mod413                                                  -0.084183   0.175227  -0.480 0.630926
diag2_mod414                                                   0.103971   0.118054   0.881 0.378476
diag2_mod424                                                  -0.005991   0.155043  -0.039 0.969177
diag2_mod425                                                   0.097361   0.134802   0.722 0.470139
diag2_mod427                                                   0.078634   0.095008   0.828 0.407866
diag2_mod428                                                   0.142541   0.089972   1.584 0.113131
diag2_mod486                                                  -0.115790   0.139813  -0.828 0.407571
diag2_mod491                                                   0.181509   0.132515   1.370 0.170773
diag2_mod496                                                   0.095009   0.103281   0.920 0.357619
diag2_mod518                                                  -0.158858   0.140608  -1.130 0.258563
diag2_mod584                                                   0.004344   0.129429   0.034 0.973228
diag2_mod585                                                   0.223211   0.115991   1.924 0.054308 .
diag2_mod599                                                  -0.033532   0.105284  -0.318 0.750115
diag2_mod682                                                   0.245589   0.129496   1.896 0.057895 .
diag2_mod707                                                   0.201487   0.114912   1.753 0.079531 .
diag2_mod780                                                   0.033005   0.134760   0.245 0.806524
diag2_modOther                                                 0.135010   0.075594   1.786 0.074102 .
diag3_mod250.02                                                0.235181   0.119012   1.976 0.048143 *
diag3_mod250.6                                                 0.574394   0.117806   4.876 1.08e-06 ***
diag3_mod272                                                  -0.092826   0.128811  -0.721 0.471136
diag3_mod276                                                   0.064008   0.076808   0.833 0.404643
diag3_mod285                                                  -0.028362   0.132320  -0.214 0.830277
diag3_mod401                                                  -0.025561   0.072086  -0.355 0.722899
diag3_mod403                                                   0.307454   0.092497   3.324 0.000888 ***
diag3_mod414                                                  -0.043974   0.091368  -0.481 0.630312
diag3_mod424                                                   0.146247   0.137588   1.063 0.287810
diag3_mod425                                                   0.121258   0.132326   0.916 0.359481
diag3_mod427                                                   0.125478   0.082141   1.528 0.126614
diag3_mod428                                                   0.096258   0.079086   1.217 0.223554
diag3_mod496                                                   0.224453   0.090286   2.486 0.012918 *
diag3_mod585                                                   0.278456   0.097547   2.855 0.004309 **
diag3_mod599                                                   0.098639   0.104477   0.944 0.345108
diag3_mod707                                                   0.163200   0.117783   1.386 0.165869
diag3_mod780                                                   0.128211   0.122852   1.044 0.296660
diag3_modOther                                                 0.119876   0.052486   2.284 0.022374 *
diag3_modV45                                                  -0.063749   0.132599  -0.481 0.630684
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 41394  on 58884  degrees of freedom
Residual deviance: 39539  on 58791  degrees of freedom
AIC: 39727

Number of Fisher Scoring iterations: 5
```