D208 - Predictive Modeling

#### **Performance Assessment - Task 2: Logistic Regression for Predictive Modeling**

#### Medical Readmission Data Set (Clean)

Joshua Trawick

Western Governors University

Table of Contents

[Part I 2](#_Toc93887135)

[A1: Research Question 2](#_Toc93887136)

[A2: Objectives and Goals 2](#_Toc93887137)

[Part II 2](#_Toc93887138)

[B1: Summary of Assumptions 2](#_Toc93887139)

[B2: Tool Benefits 2](#_Toc93887140)

[B3: Appropriate Technique 2](#_Toc93887141)

[Part III 2](#_Toc93887142)

[C1: Data Goals 2](#_Toc93887143)

[C2: Summary Statistics 2](#_Toc93887144)

[C3: Steps to Prepare the Data 2](#_Toc93887145)

[C4: Visualizations 2](#_Toc93887146)

[C5: Prepared Data Set 2](#_Toc93887147)

[Part IV 2](#_Toc93887148)

[D1: Initial Model 2](#_Toc93887149)

[D2: Justification of Model Reduction 2](#_Toc93887150)

[D3: Reduced Logistic Regression Model 2](#_Toc93887151)

[Part V 2](#_Toc93887152)

[E1: Model Comparison 2](#_Toc93887153)

[E2: Output and Calculations 2](#_Toc93887154)

[Predictions using reduced model: \*\*\*\*\*\*\*\*\*PREDICTIONS HERE\*\*\*\*\*\*\*\*\* 2](#_Toc93887155)

[E3: Code 2](#_Toc93887156)

[Part VI 2](#_Toc93887157)

[F1: Results 2](#_Toc93887158)

[F2: Recommendations 2](#_Toc93887159)

[I: Sources 2](#_Toc93887160)

## Part I

### A1: Research Question

The central research question addressed by this analysis is to determine:

What variables from the medical dataset influence a patient’s probability of being readmitted to the hospital following an initial admission (ReAdmis)?

In terms of hypothesis testing, our null hypothesis () is:

One or more variables included in the medical dataset influence readmission probability (ReAdmis) in a statistically significant way.

Additionally, our alternate hypothesis () is:

No variables included in the medical dataset influence readmission probability (ReAdmis) in a statistically significant way.

### A2: Objectives and Goals

The primary goal of the following analysis is to discover what predictors potentially increase the likelihood a patient will be readmitted after initial hospitalization (ReAdmis). This will be assessed using the programming language using the technique of multiple logistic regression to identify causal relationships between one or more predictor variables and a binary target variable.

## Part II

### B1: Summary of Assumptions

The following are basic assumptions of logistic regression:

* The assumption of binary predictor outcome (between 0 and 1)
* The assumption of linearity of continuous predictor variables
* The assumption of non-multicollinearity of predictor variables
* The assumption of no significantly influential outliers in dataset

### B2: Tool Benefits

The programming language of choice for this analysis, as previously mentioned, will be the programming language. Previously, this author has used Python to perform cleaning, transformation, and analysis. Python has been more than up to the task. However, happens to handle the process of regression analysis and model selection exceptionally well and thus became the self-evident choice for multiple logistic regression analysis. In particular, the built-in functions of the base language, as used to fit models, are incredibly simple to navigate and equally as easy to demonstrate. Additionally, the reduction process using the stepAIC() function, as another example, made the task of choosing which programming language to use for this project even easier. It is with good reason that has such a stellar reputation for handling regression models.

### B3: Appropriate Technique

Multiple logistic regression is an appropriate technique to use to accomplish our goal of finding which variables contribute to a potentially higher probability of readmission for a given patient (ReAdmis) for several reasons. Firstly, the dataset we will analyze contains 50 variables in total, each of disparate significance and utility. While some variables do not require much thought prior to elimination, others are not quite as straightforward. Therefore, running them through a multiple logistic regression model prior to discarding them is an appropriate course of action. Additionally, multiple regression will allow us to see the significance of each variable’s discrete contribution to the target variable as well as the interaction between explanatory variables themselves. Overall, regarding our objective, multiple logistic regression is more than up to the task and will adequately suit our objectives.

## Part III

### C1: Data Goals

The process we will need to complete in order to prepare the data for model selection is relatively minor, given that the raw dataset used in this project has already been cleaned in a prior project (see project D206 - Data Cleaning). Using the pre-cleaned dataset, we will first partition the data to include only those variables we intend to feed into our initial model. Because the first model selection process we will use is backward-oriented, this initial model will include all features that could possibly have a relationship to the binary target variable of readmission (ReAdmis). This will include a mix of numeric and categorical variables.

Next, we will need to ensure that the data type of each variable is appropriate for that kind of feature. For example, we will determine which variables are categorical and need to be coded as a factor in . Once the dataset for the initial model has been partitioned and transformed (or converted to the right type, at least), we will look over the dataset to ensure that we have not created any problems in the process such as silently introducing null values.

### C2: Summary Statistics

In order to get the best understanding of the selected features and their measures of central tendency, we will use an amazing library called skimr which does a phenomenal job of not only providing a great default summary view of the entire dataframe, but also allows one to customize the output. First, we will fashion a version of the skim function purpose-built for our needs here and print the output.

# Set custom skim() for C2: Summary Statistics  
# For numeric include mean, median, stdev, min, Q25, Q75, and max  
# For factor include count of unique values and value counts for each  
my\_skim <- skim\_with(  
 base = sfl(),  
 numeric = sfl(Mean = mean,  
 Median = median,  
 StDev = sd,  
 Min = min,  
 Q25 = ~ quantile(., probs = .25),  
 Q75 = ~ quantile(., probs = .75),  
 Max = max),  
 factor = sfl(Unique\_Values = n\_unique,  
 Value\_Counts = top\_counts),  
 append = FALSE  
)  
  
# Call new skim format  
my\_skim(init\_mdl)

Data summary

|  |  |
| --- | --- |
| Name | init\_mdl |
| Number of rows | 10000 |
| Number of columns | 28 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 19 |
| numeric | 9 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

|  |  |  |
| --- | --- | --- |
| skim\_variable | Unique\_Values | Value\_Counts |
| Area | 3 | Rur: 3369, Sub: 3328, Urb: 3303 |
| Marital | 5 | Wid: 2045, Mar: 2023, Sep: 1987, Nev: 1984 |
| Gender | 3 | Fem: 5018, Mal: 4768, Non: 214 |
| ReAdmis | 2 | No: 6331, Yes: 3669 |
| Soft\_drink | 2 | No: 7425, Yes: 2575 |
| Initial\_admin | 3 | Eme: 5060, Ele: 2504, Obs: 2436 |
| HighBlood | 2 | No: 5910, Yes: 4090 |
| Stroke | 2 | No: 8007, Yes: 1993 |
| Complication\_risk | 3 | Med: 4517, Hig: 3358, Low: 2125 |
| Overweight | 2 | Yes: 7094, No: 2906 |
| Arthritis | 2 | No: 6426, Yes: 3574 |
| Diabetes | 2 | No: 7262, Yes: 2738 |
| Hyperlipidemia | 2 | No: 6628, Yes: 3372 |
| BackPain | 2 | No: 5886, Yes: 4114 |
| Anxiety | 2 | No: 6785, Yes: 3215 |
| Allergic\_rhinitis | 2 | No: 6059, Yes: 3941 |
| Reflux\_esophagitis | 2 | No: 5865, Yes: 4135 |
| Asthma | 2 | No: 7107, Yes: 2893 |
| Services | 4 | Blo: 5265, Int: 3130, CT : 1225, MRI: 380 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | Mean | Median | StDev | Min | Q25 | Q75 | Max |
| Initial\_days | 34.46 | 35.84 | 26.31 | 1.00 | 7.90 | 61.16 | 71.98 |
| Children | 2.10 | 1.00 | 2.16 | 0.00 | 0.00 | 3.00 | 10.00 |
| Age | 53.51 | 53.00 | 20.64 | 18.00 | 36.00 | 71.00 | 89.00 |
| Income | 40490.50 | 33768.42 | 28521.15 | 154.08 | 19598.78 | 54296.40 | 207249.10 |
| VitD\_levels | 17.96 | 17.95 | 2.02 | 9.81 | 16.63 | 19.35 | 26.39 |
| Doc\_visits | 5.01 | 5.00 | 1.05 | 1.00 | 4.00 | 6.00 | 9.00 |
| Full\_meals\_eaten | 1.00 | 1.00 | 1.01 | 0.00 | 0.00 | 2.00 | 7.00 |
| vitD\_supp | 0.40 | 0.00 | 0.63 | 0.00 | 0.00 | 1.00 | 5.00 |
| Additional\_charges | 12934.53 | 11573.98 | 6542.60 | 3125.70 | 7986.49 | 15626.49 | 30566.07 |

The skim() output virtually speaks for itself. Our partitioned data for the initial model includes a total of 29 variables comprised of 10 numeric and 19 factor (or categorical) variables. There are a total of 10,000 rows. For the categorical variables, we have shown the names of each variable, total number of unique values for each, and the sum of each unique value for each variable respectively. Additionally, for our 10 numeric type variables, we are provided with each variable’s name followed by the mean, median, standard deviation, minimum value, lower quartile (.25), upper quartile (.75), and maximum value respectively. This summary gives us an excellent feel for the selected features for our initial model.

### C3: Steps to Prepare the Data

Now, we will begin the process of preparing the dataset, starting with loading in the necessary libraries and reading-in the cleaned dataset.

# Load in packages without messages  
suppressMessages(library(tidyverse))  
suppressMessages(library(skimr))  
suppressMessages(library(Hmisc))  
suppressMessages(library(broom))  
suppressMessages(library(olsrr))  
suppressMessages(library(car))  
suppressMessages(library(MASS))  
suppressMessages(library(InformationValue))  
  
# Read dataset in  
df <- read.csv("./data/medical\_clean.csv")  
  
# Create custom skim if not already done  
my\_skim <- skim\_with(  
 base = sfl(),  
 numeric = sfl(Mean = mean,  
 Median = median,  
 StDev = sd,  
 Min = min,  
 Q25 = ~ quantile(., probs = .25),  
 Q75 = ~ quantile(., probs = .75),  
 Max = max),  
 factor = sfl(Unique\_Values = n\_unique,  
 Value\_Counts = top\_counts),  
 append = FALSE  
)  
  
# Start with a quick skim of the data for orientation and future reference  
my\_skim(df)

Data summary

|  |  |
| --- | --- |
| Name | df |
| Number of rows | 10000 |
| Number of columns | 50 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 27 |
| numeric | 23 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| skim\_variable | min | max | empty | n\_unique | whitespace |
| Customer\_id | 6 | 7 | 0 | 10000 | 0 |
| Interaction | 36 | 36 | 0 | 10000 | 0 |
| UID | 32 | 32 | 0 | 10000 | 0 |
| City | 3 | 27 | 0 | 6072 | 0 |
| State | 2 | 2 | 0 | 52 | 0 |
| County | 3 | 21 | 0 | 1607 | 0 |
| Area | 5 | 8 | 0 | 3 | 0 |
| TimeZone | 12 | 30 | 0 | 26 | 0 |
| Job | 3 | 59 | 0 | 639 | 0 |
| Marital | 7 | 13 | 0 | 5 | 0 |
| Gender | 4 | 9 | 0 | 3 | 0 |
| ReAdmis | 2 | 3 | 0 | 2 | 0 |
| Soft\_drink | 2 | 3 | 0 | 2 | 0 |
| Initial\_admin | 18 | 21 | 0 | 3 | 0 |
| HighBlood | 2 | 3 | 0 | 2 | 0 |
| Stroke | 2 | 3 | 0 | 2 | 0 |
| Complication\_risk | 3 | 6 | 0 | 3 | 0 |
| Overweight | 2 | 3 | 0 | 2 | 0 |
| Arthritis | 2 | 3 | 0 | 2 | 0 |
| Diabetes | 2 | 3 | 0 | 2 | 0 |
| Hyperlipidemia | 2 | 3 | 0 | 2 | 0 |
| BackPain | 2 | 3 | 0 | 2 | 0 |
| Anxiety | 2 | 3 | 0 | 2 | 0 |
| Allergic\_rhinitis | 2 | 3 | 0 | 2 | 0 |
| Reflux\_esophagitis | 2 | 3 | 0 | 2 | 0 |
| Asthma | 2 | 3 | 0 | 2 | 0 |
| Services | 3 | 11 | 0 | 4 | 0 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | Mean | Median | StDev | Min | Q25 | Q75 | Max |
| CaseOrder | 5000.50 | 5000.50 | 2886.90 | 1.00 | 2500.75 | 7500.25 | 10000.00 |
| Zip | 50159.32 | 50207.00 | 27469.59 | 610.00 | 27592.00 | 72411.75 | 99929.00 |
| Lat | 38.75 | 39.42 | 5.40 | 17.97 | 35.26 | 42.04 | 70.56 |
| Lng | -91.24 | -88.40 | 15.21 | -174.21 | -97.35 | -80.44 | -65.29 |
| Population | 9965.25 | 2769.00 | 14824.76 | 0.00 | 694.75 | 13945.00 | 122814.00 |
| Children | 2.10 | 1.00 | 2.16 | 0.00 | 0.00 | 3.00 | 10.00 |
| Age | 53.51 | 53.00 | 20.64 | 18.00 | 36.00 | 71.00 | 89.00 |
| Income | 40490.50 | 33768.42 | 28521.15 | 154.08 | 19598.78 | 54296.40 | 207249.10 |
| VitD\_levels | 17.96 | 17.95 | 2.02 | 9.81 | 16.63 | 19.35 | 26.39 |
| Doc\_visits | 5.01 | 5.00 | 1.05 | 1.00 | 4.00 | 6.00 | 9.00 |
| Full\_meals\_eaten | 1.00 | 1.00 | 1.01 | 0.00 | 0.00 | 2.00 | 7.00 |
| vitD\_supp | 0.40 | 0.00 | 0.63 | 0.00 | 0.00 | 1.00 | 5.00 |
| Initial\_days | 34.46 | 35.84 | 26.31 | 1.00 | 7.90 | 61.16 | 71.98 |
| TotalCharge | 5312.17 | 5213.95 | 2180.39 | 1938.31 | 3179.37 | 7459.70 | 9180.73 |
| Additional\_charges | 12934.53 | 11573.98 | 6542.60 | 3125.70 | 7986.49 | 15626.49 | 30566.07 |
| Item1 | 3.52 | 4.00 | 1.03 | 1.00 | 3.00 | 4.00 | 8.00 |
| Item2 | 3.51 | 3.00 | 1.03 | 1.00 | 3.00 | 4.00 | 7.00 |
| Item3 | 3.51 | 4.00 | 1.03 | 1.00 | 3.00 | 4.00 | 8.00 |
| Item4 | 3.52 | 4.00 | 1.04 | 1.00 | 3.00 | 4.00 | 7.00 |
| Item5 | 3.50 | 3.00 | 1.03 | 1.00 | 3.00 | 4.00 | 7.00 |
| Item6 | 3.52 | 4.00 | 1.03 | 1.00 | 3.00 | 4.00 | 7.00 |
| Item7 | 3.49 | 3.00 | 1.02 | 1.00 | 3.00 | 4.00 | 7.00 |
| Item8 | 3.51 | 3.00 | 1.04 | 1.00 | 3.00 | 4.00 | 7.00 |

Immediately we will drop variables that, at this time, are unnecessary for our objective and keep the rest.

# Partition dataset to include only features to be initially included in model  
# Initial\_days is reordered to first position for ease of reference  
init\_mdl <- df %>%  
 dplyr::select(Initial\_days,  
 Area,  
 Children:Services,  
 Additional\_charges)  
  
# View new dataframe and assess data types  
init\_mdl %>%  
 my\_skim()

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 10000 |
| Number of columns | 28 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 19 |
| numeric | 9 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| skim\_variable | min | max | empty | n\_unique | whitespace |
| Area | 5 | 8 | 0 | 3 | 0 |
| Marital | 7 | 13 | 0 | 5 | 0 |
| Gender | 4 | 9 | 0 | 3 | 0 |
| ReAdmis | 2 | 3 | 0 | 2 | 0 |
| Soft\_drink | 2 | 3 | 0 | 2 | 0 |
| Initial\_admin | 18 | 21 | 0 | 3 | 0 |
| HighBlood | 2 | 3 | 0 | 2 | 0 |
| Stroke | 2 | 3 | 0 | 2 | 0 |
| Complication\_risk | 3 | 6 | 0 | 3 | 0 |
| Overweight | 2 | 3 | 0 | 2 | 0 |
| Arthritis | 2 | 3 | 0 | 2 | 0 |
| Diabetes | 2 | 3 | 0 | 2 | 0 |
| Hyperlipidemia | 2 | 3 | 0 | 2 | 0 |
| BackPain | 2 | 3 | 0 | 2 | 0 |
| Anxiety | 2 | 3 | 0 | 2 | 0 |
| Allergic\_rhinitis | 2 | 3 | 0 | 2 | 0 |
| Reflux\_esophagitis | 2 | 3 | 0 | 2 | 0 |
| Asthma | 2 | 3 | 0 | 2 | 0 |
| Services | 3 | 11 | 0 | 4 | 0 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | Mean | Median | StDev | Min | Q25 | Q75 | Max |
| Initial\_days | 34.46 | 35.84 | 26.31 | 1.00 | 7.90 | 61.16 | 71.98 |
| Children | 2.10 | 1.00 | 2.16 | 0.00 | 0.00 | 3.00 | 10.00 |
| Age | 53.51 | 53.00 | 20.64 | 18.00 | 36.00 | 71.00 | 89.00 |
| Income | 40490.50 | 33768.42 | 28521.15 | 154.08 | 19598.78 | 54296.40 | 207249.10 |
| VitD\_levels | 17.96 | 17.95 | 2.02 | 9.81 | 16.63 | 19.35 | 26.39 |
| Doc\_visits | 5.01 | 5.00 | 1.05 | 1.00 | 4.00 | 6.00 | 9.00 |
| Full\_meals\_eaten | 1.00 | 1.00 | 1.01 | 0.00 | 0.00 | 2.00 | 7.00 |
| vitD\_supp | 0.40 | 0.00 | 0.63 | 0.00 | 0.00 | 1.00 | 5.00 |
| Additional\_charges | 12934.53 | 11573.98 | 6542.60 | 3125.70 | 7986.49 | 15626.49 | 30566.07 |

We will next need to convert the data types of many our variables. Since there are not any variables needing type class conversion other than those which are currently of the type character, we will simply select all of those variables at once and convert them to factor variables. This will ensure that the model handles the variables as intended.

# Start with reformatting all chr variables as fct  
init\_mdl[sapply(init\_mdl,   
 is.character)] <- lapply(init\_mdl[sapply(init\_mdl,   
 is.character)],   
 as.factor)  
  
# Reassess dataframe structure using skim()  
my\_skim(init\_mdl)

Data summary

|  |  |
| --- | --- |
| Name | init\_mdl |
| Number of rows | 10000 |
| Number of columns | 28 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 19 |
| numeric | 9 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

|  |  |  |
| --- | --- | --- |
| skim\_variable | Unique\_Values | Value\_Counts |
| Area | 3 | Rur: 3369, Sub: 3328, Urb: 3303 |
| Marital | 5 | Wid: 2045, Mar: 2023, Sep: 1987, Nev: 1984 |
| Gender | 3 | Fem: 5018, Mal: 4768, Non: 214 |
| ReAdmis | 2 | No: 6331, Yes: 3669 |
| Soft\_drink | 2 | No: 7425, Yes: 2575 |
| Initial\_admin | 3 | Eme: 5060, Ele: 2504, Obs: 2436 |
| HighBlood | 2 | No: 5910, Yes: 4090 |
| Stroke | 2 | No: 8007, Yes: 1993 |
| Complication\_risk | 3 | Med: 4517, Hig: 3358, Low: 2125 |
| Overweight | 2 | Yes: 7094, No: 2906 |
| Arthritis | 2 | No: 6426, Yes: 3574 |
| Diabetes | 2 | No: 7262, Yes: 2738 |
| Hyperlipidemia | 2 | No: 6628, Yes: 3372 |
| BackPain | 2 | No: 5886, Yes: 4114 |
| Anxiety | 2 | No: 6785, Yes: 3215 |
| Allergic\_rhinitis | 2 | No: 6059, Yes: 3941 |
| Reflux\_esophagitis | 2 | No: 5865, Yes: 4135 |
| Asthma | 2 | No: 7107, Yes: 2893 |
| Services | 4 | Blo: 5265, Int: 3130, CT : 1225, MRI: 380 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | Mean | Median | StDev | Min | Q25 | Q75 | Max |
| Initial\_days | 34.46 | 35.84 | 26.31 | 1.00 | 7.90 | 61.16 | 71.98 |
| Children | 2.10 | 1.00 | 2.16 | 0.00 | 0.00 | 3.00 | 10.00 |
| Age | 53.51 | 53.00 | 20.64 | 18.00 | 36.00 | 71.00 | 89.00 |
| Income | 40490.50 | 33768.42 | 28521.15 | 154.08 | 19598.78 | 54296.40 | 207249.10 |
| VitD\_levels | 17.96 | 17.95 | 2.02 | 9.81 | 16.63 | 19.35 | 26.39 |
| Doc\_visits | 5.01 | 5.00 | 1.05 | 1.00 | 4.00 | 6.00 | 9.00 |
| Full\_meals\_eaten | 1.00 | 1.00 | 1.01 | 0.00 | 0.00 | 2.00 | 7.00 |
| vitD\_supp | 0.40 | 0.00 | 0.63 | 0.00 | 0.00 | 1.00 | 5.00 |
| Additional\_charges | 12934.53 | 11573.98 | 6542.60 | 3125.70 | 7986.49 | 15626.49 | 30566.07 |

This skim() view of the dataframe is quite useful. It shows the dataframe is comprised of two data types: factor and numeric. Thus, it appears we were successful in converting our columns to factors. The majority of the variables included are either dichotomous or otherwise nominal categorical variables. Now that we have completed our conversion process and validated the composition of our data, we may now proceed with the model selection process.

### C4: Visualizations

#### Univariate:

Now we will show all of our model’s variables using both univariate and bivariate visualizations. We’ll start with univariate histograms of the numeric variables, then we’ll look at bar charts for all of our categorical variables.

# Show histograms for all numeric variables  
par(mfrow = c(3,4))  
hist(init\_mdl %>%   
 select\_if(is.numeric))

A picture containing chart

Description automatically generated

# Partition dichotomous Yes/No variables out for plot  
dichotomous\_vars <- init\_mdl %>%   
 dplyr::select(where(~n\_distinct(.) == 2))  
  
# Show bar charts of all Yes/No variables  
dichotomous\_vars %>%  
 gather() %>%  
 count(key, value) %>%   
 ggplot(., aes(x = value, y = n)) +  
 geom\_bar(stat = "identity") +  
 facet\_wrap(~key, scales = "free", nrow = 3)

A picture containing text, crossword puzzle, scoreboard, screenshot

Description automatically generated

# Partition non-dichotomous categorical variables  
cat\_vars <- init\_mdl %>%   
 select\_if(is.factor) %>%   
 dplyr::select(where(~n\_distinct(.) != 2))  
  
# Rename levels to shorter versions to fit plots  
levels(cat\_vars$Marital) <- c("Divorced",  
 "Married",  
 "Never",  
 "Sep",  
 "Widow")  
levels(cat\_vars$Initial\_admin) <- c("Elective",  
 "Observation",  
 "Emergency")

# Create panel of bar charts for cat\_vars  
par(mfrow = c(3,2))  
barplot(table(cat\_vars$Area), main = "Geographical Area")  
barplot(table(cat\_vars$Marital), main = "Marital Status")  
barplot(table(cat\_vars$Gender), main = "Gender")  
barplot(table(cat\_vars$Initial\_admin), main = "Reason for Initial Admission")  
barplot(table(cat\_vars$Complication\_risk), main = "Complication Risk")  
barplot(table(cat\_vars$Services), main = "Services Used")

Diagram

Description automatically generated

#### Bivariate:

Now we will get a different view of the data using scatter and box plots to dive a little deeper.

# Partition model data for only numeric variables  
num\_vars <- init\_mdl %>%   
 select\_if(is.numeric)  
  
# Show scatter plot matrix of numeric variables  
pairs(num\_vars,  
 col=init\_mdl$ReAdmis)

A picture containing table

Description automatically generated

Our target variable (ReAdmis) can be identified as the red in the above visualization.

# Boxplots of our ReAdmis target variable against some numerical variables  
par(mfrow = c(2,3))  
boxplot(Initial\_days ~ ReAdmis, data = init\_mdl)  
boxplot(Children ~ ReAdmis, data = init\_mdl)  
boxplot(Age ~ ReAdmis, data = init\_mdl)  
boxplot(VitD\_levels ~ ReAdmis, data = init\_mdl)  
boxplot(Income ~ ReAdmis, data = init\_mdl)  
boxplot(Doc\_visits ~ ReAdmis, data = init\_mdl)

Box and whisker chart

Description automatically generated

### C5: Prepared Data Set

For prepared dataset, please see attached .csv file.

write.csv(init\_mdl, "./data/initmdl.csv")

## Part IV

### D1: Initial Model

The initial model will consist of essentially any potentially relevant independent variables. The initial model will then be evaluated and reduced using forward and backward stepwise selection. Considering the logistic regression equation is as follows:

and that the number of coefficients is equal to the number of independent variables, our initial model looks something like this:

Though that equation is quite unwieldy, our initial model is designed to include the maximum amount of predictor features that can reasonably be included and potentially relate to the target feature. The model also includes factor variables broken-out as dummy variables, which adds to the length of the model. In terms of our multiple logistic regression formula in , we’ve already cleaned and transformed our dataframe for the initial model, so the formula is much more concise:

ReAdmis ~ .,  
 data = init\_mdl)

### D2: Justification of Model Reduction

Our variable selection process includes the use of stepwise elimination methods, forward and backward, to systematically reduce each model one variable at a time and finally arriving at an optimal group of variables. The stepwise selection methods are similar in approach, but each can result in substantively different models. Using both forward and backward, therefore, allow us to analyze and evaluate the resulting models from each and choose the model that works best for our initial business question/hypothesis test: what relates to and/or influences a patient’s initial length of hospitalization.

Our feature selection criteria for each model reduction process will be the comparison of z-statistic p-value and AIC for individual predictors. Each step will either include or exclude the most or least useful variable respectively until no further benefit is observed. Then, the resulting reduced models will be compared against each other using the criteria, which is particularly useful for our logistic models, as well as examining the effective interpretation of coefficients. The model with the “best” combination of these measures (i.e., lowest and best explanation of coefficients) will ultimately be selected as our reduced model.

### D3: Reduced Logistic Regression Model

The reduced model is as follows:

Our multiple regression formula in , is:

ReAdmis ~ Initial\_days +  
 Children +  
 Initial\_admin +  
 HighBlood +  
 Stroke +  
 Complication\_risk +  
 Arthritis +  
 Diabetes +  
 Anxiety +  
 Allergic\_rhinitis +  
 Reflux\_esophagitis +  
 Asthma +  
 Services,  
 data = init\_mdl)

## Part V

### E1: Model Comparison

As mentioned above, the process by which the multiple regression model was reduced and selected was by using forward and backward stepwise selection. Using the MASS package, we were able to run the selection process as an iterative function and analyze the results. The initial model was overfit with very little use to begin with.

The forward selection process started by analyzing an essentially empty model, measured the impact on the z-statistic p-values and individual contribution to values for all features included in the full, initial model, and added the single feature with the most statistically significant contribution to the model at each step. This process was repeated until no further benefit was observed by adding an additional feature. This resulted in 13 statistically significant features included in the final model.

The backward elimination process used the same evaluation criterion (z-statistic p-value and individual contribution). The process is simply the reverse of forward elimination. Analysis of the initial model’s features resulted in the identification of the least statistically significant variable which was subsequently removed from the model. This step was repeated, each time removing the next-least statistically significant variable until no further benefit was observed. Backward elimination likewise concluded with 13 statistically significant features included in the optimal model.

The final output of the two models includes, perhaps most notably, their . The models were then compared by this metric and the logical explanation of coefficients. The optimal model for both selection processes resulted in identical feature selection. Typically, should forward and backward selection resulted in different models, the model with the best combination of the above ( and logical/useful coefficients) would have been selected. It’s important to note that an delta of greater than 2.0 would be sufficient to reject as inferior. (Anderson, 2002)

### E2: Output and Calculations

Consider the following summary outputs for both models including the Confusion Matrix and misclassification error rate of both.

# Summary of the forward model  
summary(stepfwd)

##   
## Call:  
## glm(formula = ReAdmis ~ Initial\_days + Initial\_admin + Services +   
## Stroke + Complication\_risk + Arthritis + Asthma + Anxiety +   
## HighBlood + Diabetes + Children + Reflux\_esophagitis + Allergic\_rhinitis,   
## family = binomial, data = init\_mdl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.5137 0.0000 0.0000 0.0001 3.3847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -80.77379 4.51401 -17.894 < 2e-16 \*\*\*  
## Initial\_days 1.47447 0.08229 17.918 < 2e-16 \*\*\*  
## Initial\_adminEmergency Admission 2.52225 0.27709 9.103 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 0.77122 0.27467 2.808 0.00499 \*\*   
## ServicesCT Scan 1.58311 0.35925 4.407 1.05e-05 \*\*\*  
## ServicesIntravenous 0.01598 0.22706 0.070 0.94391   
## ServicesMRI 2.74402 0.48841 5.618 1.93e-08 \*\*\*  
## StrokeYes 1.62981 0.26347 6.186 6.17e-10 \*\*\*  
## Complication\_riskLow -1.85203 0.29084 -6.368 1.92e-10 \*\*\*  
## Complication\_riskMedium -0.33849 0.22983 -1.473 0.14081   
## ArthritisYes -1.34063 0.22254 -6.024 1.70e-09 \*\*\*  
## AsthmaYes -1.34759 0.23301 -5.783 7.32e-09 \*\*\*  
## AnxietyYes -1.02489 0.22300 -4.596 4.31e-06 \*\*\*  
## HighBloodYes 0.86692 0.21263 4.077 4.56e-05 \*\*\*  
## DiabetesYes 0.46011 0.22806 2.018 0.04364 \*   
## Children 0.09512 0.04528 2.101 0.03566 \*   
## Reflux\_esophagitisYes -0.40251 0.20678 -1.947 0.05158 .   
## Allergic\_rhinitisYes -0.33276 0.20464 -1.626 0.10394   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 13145.70 on 9999 degrees of freedom  
## Residual deviance: 668.51 on 9982 degrees of freedom  
## AIC: 704.51  
##   
## Number of Fisher Scoring iterations: 12

# Create predicted values based on stepfwd  
predicted <- predict(stepfwd,  
 init\_mdl,  
 type = "response")  
  
# Convert target var (ReAdmis) to binary 1/0 vals  
predmdl <- init\_mdl  
predmdl$ReAdmis <- ifelse(predmdl$ReAdmis == "Yes",  
 1,  
 0)  
  
# Compute optimal prob cutoff score  
optimal <- optimalCutoff(predmdl$ReAdmis,  
 predicted)[1]  
  
# Create confusion matrix of actuals and predicted values  
as.matrix(confusionMatrix(predmdl$ReAdmis,  
 predicted))

## 0 1  
## 0 6256 63  
## 1 75 3606

# Declare misclassification error  
misClassError(predmdl$ReAdmis,  
 predicted,  
 threshold=optimal)

## [1] 0.0133

# Summary of the backward elimination model  
summary(stepbck)

##   
## Call:  
## glm(formula = ReAdmis ~ Initial\_days + Children + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Arthritis + Diabetes +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services, family = binomial, data = init\_mdl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.5137 0.0000 0.0000 0.0001 3.3847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -80.77379 4.51401 -17.894 < 2e-16 \*\*\*  
## Initial\_days 1.47447 0.08229 17.918 < 2e-16 \*\*\*  
## Children 0.09512 0.04528 2.101 0.03566 \*   
## Initial\_adminEmergency Admission 2.52225 0.27709 9.103 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 0.77122 0.27467 2.808 0.00499 \*\*   
## HighBloodYes 0.86692 0.21263 4.077 4.56e-05 \*\*\*  
## StrokeYes 1.62981 0.26347 6.186 6.17e-10 \*\*\*  
## Complication\_riskLow -1.85203 0.29084 -6.368 1.92e-10 \*\*\*  
## Complication\_riskMedium -0.33849 0.22983 -1.473 0.14081   
## ArthritisYes -1.34063 0.22254 -6.024 1.70e-09 \*\*\*  
## DiabetesYes 0.46011 0.22806 2.018 0.04364 \*   
## AnxietyYes -1.02489 0.22300 -4.596 4.31e-06 \*\*\*  
## Allergic\_rhinitisYes -0.33276 0.20464 -1.626 0.10394   
## Reflux\_esophagitisYes -0.40251 0.20678 -1.947 0.05158 .   
## AsthmaYes -1.34759 0.23301 -5.783 7.32e-09 \*\*\*  
## ServicesCT Scan 1.58311 0.35925 4.407 1.05e-05 \*\*\*  
## ServicesIntravenous 0.01598 0.22706 0.070 0.94391   
## ServicesMRI 2.74402 0.48841 5.618 1.93e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 13145.70 on 9999 degrees of freedom  
## Residual deviance: 668.51 on 9982 degrees of freedom  
## AIC: 704.51  
##   
## Number of Fisher Scoring iterations: 12

# Create predicted values based on stepbck  
predicted <- predict(stepbck,  
 init\_mdl,  
 type = "response")  
  
# Convert target var (ReAdmis) to binary 1/0 vals  
predmdl <- init\_mdl  
predmdl$ReAdmis <- ifelse(predmdl$ReAdmis == "Yes",  
 1,  
 0)  
  
# Compute optimal prob cutoff score  
optimal <- optimalCutoff(predmdl$ReAdmis,  
 predicted)[1]  
  
# Create confusion matrix of actuals and predicted values  
as.matrix(confusionMatrix(predmdl$ReAdmis,  
 predicted))

## 0 1  
## 0 6256 63  
## 1 75 3606

# Declare misclassification error  
misClassError(predmdl$ReAdmis,  
 predicted,  
 threshold=optimal)

## [1] 0.0133

### Predictions using reduced model:

# Create a dataframe of 20 predictor vectors  
p\_Initial\_days <- seq(min(init\_mdl$Initial\_days),  
 max(init\_mdl$Initial\_days),  
 by = (max(init\_mdl$Initial\_days) - min(init\_mdl$Initial\_days)) / 19)  
p\_Initial\_admin <- rep\_len(levels(init\_mdl$Initial\_admin),  
 length.out = 20)  
p\_Services <- rep\_len(levels(init\_mdl$Services),  
 length.out = 20)  
p\_Stroke <- rep\_len(levels(init\_mdl$Stroke),  
 length.out = 20)  
p\_Complication\_risk <- rep\_len(levels(init\_mdl$Complication\_risk),  
 length.out = 20)  
p\_Arthritis <- rep\_len(levels(init\_mdl$Arthritis),  
 length.out = 20)  
p\_Asthma <- rep\_len(levels(init\_mdl$Asthma),  
 length.out = 20)  
p\_Anxiety <- rep\_len(levels(init\_mdl$Anxiety),  
 length.out = 20)  
p\_HighBlood <- rep\_len(levels(init\_mdl$HighBlood),  
 length.out = 20)  
p\_Diabetes <- rep\_len(levels(init\_mdl$Diabetes),  
 length.out = 20)  
p\_Children <- sample(init\_mdl$Children,  
 20,  
 replace = TRUE)  
p\_Reflux\_esophagitis <- rep\_len(levels(init\_mdl$Reflux\_esophagitis),  
 length.out = 20)  
p\_Allergic\_rhinitis <- rep\_len(levels(init\_mdl$Allergic\_rhinitis),  
 length.out = 20)  
  
p\_mdl <- data.frame(Initial\_days = p\_Initial\_days,  
 Initial\_admin = p\_Initial\_admin,  
 Services = p\_Services,  
 Stroke = p\_Stroke,  
 Complication\_risk = p\_Complication\_risk,  
 Arthritis = p\_Arthritis,  
 Asthma = p\_Asthma,  
 Anxiety = p\_Anxiety,  
 HighBlood = p\_HighBlood,  
 Diabetes = p\_Diabetes,  
 Children = p\_Children,  
 Reflux\_esophagitis = p\_Reflux\_esophagitis,  
 Allergic\_rhinitis = p\_Allergic\_rhinitis)  
  
# Add predictions column p\_ReAdmis  
p\_mdl$p\_ReAdmis <- predict(stepfwd,  
 new = p\_mdl,  
 type = "response")  
  
p\_mdl

## Initial\_days Initial\_admin Services Stroke Complication\_risk  
## 1 1.001981 Elective Admission Blood Work No High  
## 2 4.737745 Emergency Admission CT Scan Yes Low  
## 3 8.473508 Observation Admission Intravenous No Medium  
## 4 12.209272 Elective Admission MRI Yes High  
## 5 15.945035 Emergency Admission Blood Work No Low  
## 6 19.680799 Observation Admission CT Scan Yes Medium  
## 7 23.416563 Elective Admission Intravenous No High  
## 8 27.152326 Emergency Admission MRI Yes Low  
## 9 30.888090 Observation Admission Blood Work No Medium  
## 10 34.623854 Elective Admission CT Scan Yes High  
## 11 38.359617 Emergency Admission Intravenous No Low  
## 12 42.095381 Observation Admission MRI Yes Medium  
## 13 45.831145 Elective Admission Blood Work No High  
## 14 49.566908 Emergency Admission CT Scan Yes Low  
## 15 53.302672 Observation Admission Intravenous No Medium  
## 16 57.038435 Elective Admission MRI Yes High  
## 17 60.774199 Emergency Admission Blood Work No Low  
## 18 64.509963 Observation Admission CT Scan Yes Medium  
## 19 68.245726 Elective Admission Intravenous No High  
## 20 71.981490 Emergency Admission MRI Yes Low  
## Arthritis Asthma Anxiety HighBlood Diabetes Children Reflux\_esophagitis  
## 1 No No No No No 2 No  
## 2 Yes Yes Yes Yes Yes 3 Yes  
## 3 No No No No No 4 No  
## 4 Yes Yes Yes Yes Yes 4 Yes  
## 5 No No No No No 0 No  
## 6 Yes Yes Yes Yes Yes 1 Yes  
## 7 No No No No No 3 No  
## 8 Yes Yes Yes Yes Yes 7 Yes  
## 9 No No No No No 3 No  
## 10 Yes Yes Yes Yes Yes 1 Yes  
## 11 No No No No No 3 No  
## 12 Yes Yes Yes Yes Yes 0 Yes  
## 13 No No No No No 2 No  
## 14 Yes Yes Yes Yes Yes 1 Yes  
## 15 No No No No No 0 No  
## 16 Yes Yes Yes Yes Yes 1 Yes  
## 17 No No No No No 4 No  
## 18 Yes Yes Yes Yes Yes 3 Yes  
## 19 No No No No No 1 No  
## 20 Yes Yes Yes Yes Yes 7 Yes  
## Allergic\_rhinitis p\_ReAdmis  
## 1 No 2.220446e-16  
## 2 Yes 2.220446e-16  
## 3 No 2.220446e-16  
## 4 Yes 2.220446e-16  
## 5 No 2.220446e-16  
## 6 Yes 2.220446e-16  
## 7 No 2.220446e-16  
## 8 Yes 2.220446e-16  
## 9 No 2.220446e-16  
## 10 Yes 1.489402e-13  
## 11 No 8.055425e-11  
## 12 Yes 4.057292e-08  
## 13 No 2.244859e-06  
## 14 Yes 1.077629e-03  
## 15 No 1.503509e-01  
## 16 Yes 9.907635e-01  
## 17 No 9.999492e-01  
## 18 Yes 9.999997e-01  
## 19 No 1.000000e+00  
## 20 Yes 1.000000e+00

### E3: Code

# Initial/upper logit model  
logitfull <- glm(ReAdmis ~ .,  
 data = init\_mdl,  
 family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

# Lower logit model  
logitlow <- glm(ReAdmis ~ 1,  
 data = init\_mdl,  
 family = binomial)

# Forward stepwise model selection (suppress perfect separation warnings)  
stepfwd <- suppressWarnings(stepAIC(logitlow,  
 scope = list(lower = logitlow,  
 upper = logitfull)))

## Start: AIC=13147.7  
## ReAdmis ~ 1  
##   
## Df Deviance AIC  
## + Initial\_days 1 978.7 982.7  
## + Children 1 13140.2 13144.2  
## + Services 3 13136.9 13144.9  
## + Asthma 1 13142.8 13146.8  
## + Age 1 13143.2 13147.2  
## <none> 13145.7 13147.7  
## + Initial\_admin 2 13141.8 13147.8  
## + Additional\_charges 1 13143.9 13147.9  
## + BackPain 1 13143.9 13147.9  
## + Full\_meals\_eaten 1 13144.2 13148.2  
## + Income 1 13144.4 13148.4  
## + vitD\_supp 1 13144.5 13148.5  
## + Overweight 1 13145.0 13149.0  
## + Soft\_drink 1 13145.1 13149.1  
## + Arthritis 1 13145.1 13149.1  
## + Reflux\_esophagitis 1 13145.4 13149.4  
## + Allergic\_rhinitis 1 13145.5 13149.5  
## + Hyperlipidemia 1 13145.5 13149.5  
## + VitD\_levels 1 13145.5 13149.5  
## + Diabetes 1 13145.6 13149.6  
## + Anxiety 1 13145.6 13149.6  
## + HighBlood 1 13145.7 13149.7  
## + Stroke 1 13145.7 13149.7  
## + Doc\_visits 1 13145.7 13149.7  
## + Gender 2 13144.1 13150.1  
## + Marital 4 13140.6 13150.6  
## + Area 2 13145.0 13151.0  
## + Complication\_risk 2 13145.5 13151.5  
##   
## Step: AIC=982.7  
## ReAdmis ~ Initial\_days  
##   
## Df Deviance AIC  
## + Initial\_admin 2 899.3 907.3  
## + Stroke 1 943.9 949.9  
## + Services 3 943.5 953.5  
## + Complication\_risk 2 948.1 956.1  
## + Asthma 1 951.5 957.5  
## + Arthritis 1 959.8 965.8  
## + Anxiety 1 965.4 971.4  
## + HighBlood 1 967.3 973.3  
## + Additional\_charges 1 968.2 974.2  
## + Hyperlipidemia 1 975.3 981.3  
## + Children 1 975.8 981.8  
## + BackPain 1 976.4 982.4  
## <none> 978.7 982.7  
## + Diabetes 1 977.1 983.1  
## + Allergic\_rhinitis 1 977.2 983.2  
## + Reflux\_esophagitis 1 977.7 983.7  
## + VitD\_levels 1 978.1 984.1  
## + Full\_meals\_eaten 1 978.3 984.3  
## + Soft\_drink 1 978.3 984.3  
## + Overweight 1 978.4 984.4  
## + vitD\_supp 1 978.4 984.4  
## + Income 1 978.6 984.6  
## + Doc\_visits 1 978.6 984.6  
## + Age 1 978.7 984.7  
## + Area 2 977.6 985.6  
## + Gender 2 978.1 986.1  
## + Marital 4 975.1 987.1  
## - Initial\_days 1 13145.7 13147.7  
##   
## Step: AIC=907.3  
## ReAdmis ~ Initial\_days + Initial\_admin  
##   
## Df Deviance AIC  
## + Services 3 856.6 870.6  
## + Stroke 1 864.3 874.3  
## + Complication\_risk 2 867.8 879.8  
## + Asthma 1 871.0 881.0  
## + Arthritis 1 872.2 882.2  
## + Anxiety 1 885.4 895.4  
## + HighBlood 1 887.4 897.4  
## + Additional\_charges 1 890.0 900.0  
## + Children 1 896.0 906.0  
## + Hyperlipidemia 1 897.1 907.1  
## <none> 899.3 907.3  
## + Allergic\_rhinitis 1 897.4 907.4  
## + Reflux\_esophagitis 1 897.5 907.5  
## + BackPain 1 897.5 907.5  
## + Overweight 1 898.3 908.3  
## + Diabetes 1 898.5 908.5  
## + VitD\_levels 1 898.6 908.6  
## + Soft\_drink 1 898.8 908.8  
## + Full\_meals\_eaten 1 899.1 909.1  
## + vitD\_supp 1 899.2 909.2  
## + Doc\_visits 1 899.3 909.3  
## + Age 1 899.3 909.3  
## + Income 1 899.3 909.3  
## + Gender 2 897.6 909.6  
## + Area 2 898.7 910.7  
## + Marital 4 897.5 913.5  
## - Initial\_admin 2 978.7 982.7  
## - Initial\_days 1 13141.8 13147.8  
##   
## Step: AIC=870.64  
## ReAdmis ~ Initial\_days + Initial\_admin + Services  
##   
## Df Deviance AIC  
## + Stroke 1 821.2 837.2  
## + Complication\_risk 2 821.3 839.3  
## + Asthma 1 827.5 843.5  
## + Arthritis 1 829.0 845.0  
## + Anxiety 1 842.4 858.4  
## + HighBlood 1 844.0 860.0  
## + Additional\_charges 1 848.1 864.1  
## + Children 1 852.7 868.7  
## + Reflux\_esophagitis 1 854.6 870.6  
## <none> 856.6 870.6  
## + Hyperlipidemia 1 854.6 870.6  
## + Allergic\_rhinitis 1 854.9 870.9  
## + BackPain 1 855.4 871.4  
## + Diabetes 1 856.1 872.1  
## + VitD\_levels 1 856.1 872.1  
## + Overweight 1 856.1 872.1  
## + Soft\_drink 1 856.4 872.4  
## + vitD\_supp 1 856.4 872.4  
## + Gender 2 854.4 872.4  
## + Income 1 856.5 872.5  
## + Full\_meals\_eaten 1 856.5 872.5  
## + Doc\_visits 1 856.6 872.6  
## + Age 1 856.6 872.6  
## + Area 2 856.5 874.5  
## + Marital 4 853.5 875.5  
## - Services 3 899.3 907.3  
## - Initial\_admin 2 943.5 953.5  
## - Initial\_days 1 13133.1 13145.1  
##   
## Step: AIC=837.22  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke  
##   
## Df Deviance AIC  
## + Complication\_risk 2 784.3 804.3  
## + Arthritis 1 793.0 811.0  
## + Asthma 1 793.6 811.6  
## + Anxiety 1 805.6 823.6  
## + HighBlood 1 808.7 826.7  
## + Additional\_charges 1 814.2 832.2  
## + Children 1 816.2 834.2  
## + Hyperlipidemia 1 818.5 836.5  
## + Reflux\_esophagitis 1 819.1 837.1  
## <none> 821.2 837.2  
## + Allergic\_rhinitis 1 819.6 837.6  
## + BackPain 1 819.7 837.7  
## + Diabetes 1 819.9 837.9  
## + VitD\_levels 1 820.6 838.6  
## + Overweight 1 820.8 838.8  
## + vitD\_supp 1 820.9 838.9  
## + Soft\_drink 1 820.9 838.9  
## + Full\_meals\_eaten 1 821.1 839.1  
## + Income 1 821.2 839.2  
## + Age 1 821.2 839.2  
## + Doc\_visits 1 821.2 839.2  
## + Gender 2 819.6 839.6  
## + Area 2 820.8 840.8  
## + Marital 4 818.6 842.6  
## - Stroke 1 856.6 870.6  
## - Services 3 864.3 874.3  
## - Initial\_admin 2 907.7 919.7  
## - Initial\_days 1 13133.1 13147.1  
##   
## Step: AIC=804.27  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk  
##   
## Df Deviance AIC  
## + Arthritis 1 751.9 773.9  
## + Asthma 1 752.4 774.4  
## + Anxiety 1 766.5 788.5  
## + HighBlood 1 773.0 795.0  
## + Additional\_charges 1 778.4 800.4  
## + Children 1 779.6 801.6  
## + Reflux\_esophagitis 1 782.2 804.2  
## <none> 784.3 804.3  
## + Allergic\_rhinitis 1 782.6 804.6  
## + Diabetes 1 782.8 804.8  
## + Hyperlipidemia 1 782.8 804.8  
## + BackPain 1 782.9 804.9  
## + Overweight 1 783.2 805.2  
## + Soft\_drink 1 783.4 805.4  
## + vitD\_supp 1 783.7 805.7  
## + VitD\_levels 1 784.1 806.1  
## + Doc\_visits 1 784.2 806.2  
## + Age 1 784.3 806.3  
## + Full\_meals\_eaten 1 784.3 806.3  
## + Income 1 784.3 806.3  
## + Gender 2 783.1 807.1  
## + Area 2 784.1 808.1  
## + Marital 4 781.7 809.7  
## - Complication\_risk 2 821.2 837.2  
## - Stroke 1 821.3 839.3  
## - Services 3 832.0 846.0  
## - Initial\_admin 2 874.6 890.6  
## - Initial\_days 1 13132.9 13150.9  
##   
## Step: AIC=773.88  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis  
##   
## Df Deviance AIC  
## + Asthma 1 718.4 742.4  
## + Anxiety 1 733.0 757.0  
## + HighBlood 1 738.8 762.8  
## + Additional\_charges 1 744.7 768.7  
## + Children 1 747.1 771.1  
## + Reflux\_esophagitis 1 749.2 773.2  
## + Diabetes 1 749.4 773.4  
## <none> 751.9 773.9  
## + Hyperlipidemia 1 750.3 774.3  
## + Allergic\_rhinitis 1 750.4 774.4  
## + Soft\_drink 1 750.7 774.7  
## + BackPain 1 751.0 775.0  
## + Overweight 1 751.3 775.3  
## + VitD\_levels 1 751.5 775.5  
## + vitD\_supp 1 751.5 775.5  
## + Doc\_visits 1 751.8 775.8  
## + Full\_meals\_eaten 1 751.9 775.9  
## + Age 1 751.9 775.9  
## + Income 1 751.9 775.9  
## + Gender 2 750.4 776.4  
## + Area 2 751.7 777.7  
## + Marital 4 749.5 779.5  
## - Arthritis 1 784.3 804.3  
## - Stroke 1 790.7 810.7  
## - Complication\_risk 2 793.0 811.0  
## - Services 3 800.4 816.4  
## - Initial\_admin 2 850.9 868.9  
## - Initial\_days 1 13132.3 13152.3  
##   
## Step: AIC=742.35  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma  
##   
## Df Deviance AIC  
## + Anxiety 1 699.3 725.3  
## + HighBlood 1 704.2 730.2  
## + Additional\_charges 1 710.6 736.6  
## + Children 1 713.8 739.8  
## + Diabetes 1 715.5 741.5  
## + Reflux\_esophagitis 1 715.7 741.7  
## <none> 718.4 742.4  
## + Hyperlipidemia 1 716.5 742.5  
## + Soft\_drink 1 716.7 742.7  
## + Allergic\_rhinitis 1 716.8 742.8  
## + BackPain 1 717.3 743.3  
## + Overweight 1 717.8 743.8  
## + VitD\_levels 1 717.9 743.9  
## + vitD\_supp 1 718.0 744.0  
## + Doc\_visits 1 718.2 744.2  
## + Full\_meals\_eaten 1 718.2 744.2  
## + Income 1 718.3 744.3  
## + Age 1 718.3 744.3  
## + Gender 2 716.5 744.5  
## + Area 2 717.9 745.9  
## + Marital 4 716.0 748.0  
## - Asthma 1 751.9 773.9  
## - Arthritis 1 752.4 774.4  
## - Stroke 1 755.7 777.7  
## - Complication\_risk 2 764.6 784.6  
## - Services 3 768.2 786.2  
## - Initial\_admin 2 822.0 842.0  
## - Initial\_days 1 13129.3 13151.3  
##   
## Step: AIC=725.27  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety  
##   
## Df Deviance AIC  
## + HighBlood 1 683.2 711.2  
## + Additional\_charges 1 689.8 717.8  
## + Children 1 693.8 721.8  
## + Reflux\_esophagitis 1 696.2 724.2  
## + Soft\_drink 1 697.0 725.0  
## + Diabetes 1 697.1 725.1  
## + Allergic\_rhinitis 1 697.2 725.2  
## <none> 699.3 725.3  
## + BackPain 1 697.3 725.3  
## + Hyperlipidemia 1 697.6 725.6  
## + VitD\_levels 1 698.6 726.6  
## + vitD\_supp 1 698.6 726.6  
## + Overweight 1 698.7 726.7  
## + Age 1 699.1 727.1  
## + Full\_meals\_eaten 1 699.1 727.1  
## + Doc\_visits 1 699.1 727.1  
## + Income 1 699.1 727.1  
## + Gender 2 698.1 728.1  
## + Area 2 699.0 729.0  
## + Marital 4 696.5 730.5  
## - Anxiety 1 718.4 742.4  
## - Asthma 1 733.0 757.0  
## - Arthritis 1 734.3 758.3  
## - Stroke 1 738.0 762.0  
## - Services 3 749.6 769.6  
## - Complication\_risk 2 747.9 769.9  
## - Initial\_admin 2 806.8 828.8  
## - Initial\_days 1 13129.2 13153.2  
##   
## Step: AIC=711.18  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety + HighBlood  
##   
## Df Deviance AIC  
## + Diabetes 1 678.7 708.7  
## + Children 1 679.1 709.1  
## + Reflux\_esophagitis 1 679.8 709.8  
## + Allergic\_rhinitis 1 680.8 710.8  
## + Soft\_drink 1 680.9 710.9  
## <none> 683.2 711.2  
## + Hyperlipidemia 1 681.5 711.5  
## + BackPain 1 681.5 711.5  
## + Overweight 1 682.0 712.0  
## + VitD\_levels 1 682.6 712.6  
## + vitD\_supp 1 682.7 712.7  
## + Additional\_charges 1 682.8 712.8  
## + Full\_meals\_eaten 1 682.8 712.8  
## + Income 1 683.0 713.0  
## + Age 1 683.1 713.1  
## + Doc\_visits 1 683.2 713.2  
## + Gender 2 681.6 713.6  
## + Area 2 683.0 715.0  
## + Marital 4 681.0 717.0  
## - HighBlood 1 699.3 725.3  
## - Anxiety 1 704.2 730.2  
## - Asthma 1 718.6 744.6  
## - Arthritis 1 720.5 746.5  
## - Stroke 1 722.7 748.7  
## - Complication\_risk 2 730.6 754.6  
## - Services 3 733.8 755.8  
## - Initial\_admin 2 792.7 816.7  
## - Initial\_days 1 13129.2 13155.2  
##   
## Step: AIC=708.7  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety + HighBlood +   
## Diabetes  
##   
## Df Deviance AIC  
## + Children 1 674.7 706.7  
## + Reflux\_esophagitis 1 675.7 707.7  
## + Allergic\_rhinitis 1 676.2 708.2  
## + Soft\_drink 1 676.7 708.7  
## <none> 678.7 708.7  
## + Hyperlipidemia 1 677.0 709.0  
## + BackPain 1 677.1 709.1  
## + Overweight 1 677.6 709.6  
## + VitD\_levels 1 678.1 710.1  
## + Additional\_charges 1 678.2 710.2  
## + vitD\_supp 1 678.3 710.3  
## + Income 1 678.4 710.4  
## + Full\_meals\_eaten 1 678.4 710.4  
## + Age 1 678.5 710.5  
## + Doc\_visits 1 678.7 710.7  
## + Gender 2 677.1 711.1  
## - Diabetes 1 683.2 711.2  
## + Area 2 678.6 712.6  
## + Marital 4 676.1 714.1  
## - HighBlood 1 697.1 725.1  
## - Anxiety 1 699.0 727.0  
## - Asthma 1 714.8 742.8  
## - Arthritis 1 717.7 745.7  
## - Stroke 1 720.4 748.4  
## - Complication\_risk 2 726.7 752.7  
## - Services 3 728.8 752.8  
## - Initial\_admin 2 788.5 814.5  
## - Initial\_days 1 13129.1 13157.1  
##   
## Step: AIC=706.71  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety + HighBlood +   
## Diabetes + Children  
##   
## Df Deviance AIC  
## + Reflux\_esophagitis 1 671.2 705.2  
## + Allergic\_rhinitis 1 672.3 706.3  
## + Soft\_drink 1 672.7 706.7  
## <none> 674.7 706.7  
## + Hyperlipidemia 1 672.8 706.8  
## + BackPain 1 673.1 707.1  
## + Overweight 1 673.8 707.8  
## + Additional\_charges 1 674.1 708.1  
## + VitD\_levels 1 674.3 708.3  
## + vitD\_supp 1 674.4 708.4  
## + Age 1 674.4 708.4  
## + Full\_meals\_eaten 1 674.5 708.5  
## + Income 1 674.5 708.5  
## - Children 1 678.7 708.7  
## + Doc\_visits 1 674.7 708.7  
## - Diabetes 1 679.1 709.1  
## + Gender 2 673.5 709.5  
## + Area 2 674.6 710.6  
## + Marital 4 671.8 711.8  
## - HighBlood 1 691.5 721.5  
## - Anxiety 1 695.5 725.5  
## - Asthma 1 710.3 740.3  
## - Arthritis 1 713.8 743.8  
## - Stroke 1 716.7 746.7  
## - Complication\_risk 2 722.7 750.7  
## - Services 3 725.1 751.1  
## - Initial\_admin 2 785.3 813.3  
## - Initial\_days 1 13123.5 13153.5  
##   
## Step: AIC=705.17  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety + HighBlood +   
## Diabetes + Children + Reflux\_esophagitis  
##   
## Df Deviance AIC  
## + Allergic\_rhinitis 1 668.5 704.5  
## <none> 671.2 705.2  
## + Soft\_drink 1 669.2 705.2  
## + BackPain 1 669.4 705.4  
## + Hyperlipidemia 1 669.4 705.4  
## + Overweight 1 670.0 706.0  
## - Reflux\_esophagitis 1 674.7 706.7  
## + Additional\_charges 1 670.8 706.8  
## + VitD\_levels 1 670.8 706.8  
## + Full\_meals\_eaten 1 670.8 706.8  
## + Income 1 670.9 706.9  
## + vitD\_supp 1 670.9 706.9  
## + Age 1 671.0 707.0  
## + Doc\_visits 1 671.2 707.2  
## - Diabetes 1 675.2 707.2  
## - Children 1 675.7 707.7  
## + Gender 2 670.2 708.2  
## + Area 2 671.1 709.1  
## + Marital 4 668.0 710.0  
## - HighBlood 1 688.2 720.2  
## - Anxiety 1 692.6 724.6  
## - Asthma 1 707.0 739.0  
## - Arthritis 1 710.7 742.7  
## - Stroke 1 713.4 745.4  
## - Complication\_risk 2 719.1 749.1  
## - Services 3 722.0 750.0  
## - Initial\_admin 2 783.5 813.5  
## - Initial\_days 1 13123.3 13155.3  
##   
## Step: AIC=704.51  
## ReAdmis ~ Initial\_days + Initial\_admin + Services + Stroke +   
## Complication\_risk + Arthritis + Asthma + Anxiety + HighBlood +   
## Diabetes + Children + Reflux\_esophagitis + Allergic\_rhinitis  
##   
## Df Deviance AIC  
## <none> 668.5 704.5  
## + BackPain 1 666.6 704.6  
## + Soft\_drink 1 666.8 704.8  
## + Hyperlipidemia 1 667.0 705.0  
## + Overweight 1 667.1 705.1  
## - Allergic\_rhinitis 1 671.2 705.2  
## + Full\_meals\_eaten 1 668.2 706.2  
## + vitD\_supp 1 668.2 706.2  
## + Additional\_charges 1 668.3 706.3  
## + VitD\_levels 1 668.3 706.3  
## + Income 1 668.3 706.3  
## - Reflux\_esophagitis 1 672.3 706.3  
## + Age 1 668.4 706.4  
## + Doc\_visits 1 668.5 706.5  
## - Diabetes 1 672.6 706.6  
## - Children 1 673.0 707.0  
## + Gender 2 667.6 707.6  
## + Area 2 668.4 708.4  
## + Marital 4 665.3 709.3  
## - HighBlood 1 685.8 719.8  
## - Anxiety 1 690.8 724.8  
## - Asthma 1 704.7 738.7  
## - Arthritis 1 708.4 742.4  
## - Stroke 1 710.7 744.7  
## - Complication\_risk 2 716.5 748.5  
## - Services 3 719.5 749.5  
## - Initial\_admin 2 782.4 814.4  
## - Initial\_days 1 13123.1 13157.1

# Backward stepwise model selection (suppress perfect separation warnings)  
stepbck <- suppressWarnings(stepAIC(logitfull,  
 direction = "backward"))

## Start: AIC=730.28  
## ReAdmis ~ Initial\_days + Area + Children + Age + Income + Marital +   
## Gender + VitD\_levels + Doc\_visits + Full\_meals\_eaten + vitD\_supp +   
## Soft\_drink + Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Marital 4 659.2 725.2  
## - Area 2 656.5 726.5  
## - Gender 2 657.1 727.1  
## - Doc\_visits 1 656.3 728.3  
## - Income 1 656.4 728.4  
## - Full\_meals\_eaten 1 656.5 728.5  
## - VitD\_levels 1 656.5 728.5  
## - Age 1 656.5 728.5  
## - vitD\_supp 1 656.6 728.6  
## - Additional\_charges 1 656.8 728.8  
## - HighBlood 1 657.0 729.0  
## - BackPain 1 657.5 729.5  
## - Hyperlipidemia 1 657.7 729.7  
## - Soft\_drink 1 658.1 730.1  
## - Overweight 1 658.2 730.2  
## <none> 656.3 730.3  
## - Allergic\_rhinitis 1 658.5 730.5  
## - Children 1 660.0 732.0  
## - Reflux\_esophagitis 1 660.4 732.4  
## - Diabetes 1 660.8 732.8  
## - Anxiety 1 679.2 751.2  
## - Asthma 1 694.4 766.4  
## - Arthritis 1 696.4 768.4  
## - Stroke 1 697.2 769.2  
## - Complication\_risk 2 700.9 770.9  
## - Services 3 706.3 774.3  
## - Initial\_admin 2 761.4 831.4  
## - Initial\_days 1 13105.9 13177.9  
##   
## Step: AIC=725.16  
## ReAdmis ~ Initial\_days + Area + Children + Age + Income + Gender +   
## VitD\_levels + Doc\_visits + Full\_meals\_eaten + vitD\_supp +   
## Soft\_drink + Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Area 2 659.4 721.4  
## - Gender 2 660.1 722.1  
## - Doc\_visits 1 659.2 723.2  
## - Income 1 659.2 723.2  
## - VitD\_levels 1 659.4 723.4  
## - Age 1 659.4 723.4  
## - Full\_meals\_eaten 1 659.4 723.4  
## - vitD\_supp 1 659.5 723.5  
## - Additional\_charges 1 659.6 723.6  
## - HighBlood 1 660.0 724.0  
## - Hyperlipidemia 1 660.7 724.7  
## - BackPain 1 660.8 724.8  
## - Soft\_drink 1 660.9 724.9  
## - Overweight 1 661.0 725.0  
## <none> 659.2 725.2  
## - Allergic\_rhinitis 1 661.4 725.4  
## - Children 1 662.6 726.6  
## - Reflux\_esophagitis 1 663.0 727.0  
## - Diabetes 1 663.1 727.1  
## - Anxiety 1 681.8 745.8  
## - Asthma 1 697.0 761.0  
## - Arthritis 1 699.1 763.1  
## - Stroke 1 699.8 763.8  
## - Complication\_risk 2 703.9 765.9  
## - Services 3 708.5 768.5  
## - Initial\_admin 2 764.5 826.5  
## - Initial\_days 1 13111.3 13175.3  
##   
## Step: AIC=721.41  
## ReAdmis ~ Initial\_days + Children + Age + Income + Gender + VitD\_levels +   
## Doc\_visits + Full\_meals\_eaten + vitD\_supp + Soft\_drink +   
## Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Gender 2 660.3 718.3  
## - Doc\_visits 1 659.4 719.4  
## - Income 1 659.5 719.5  
## - VitD\_levels 1 659.6 719.6  
## - Full\_meals\_eaten 1 659.7 719.7  
## - Age 1 659.7 719.7  
## - vitD\_supp 1 659.8 719.8  
## - Additional\_charges 1 659.9 719.9  
## - HighBlood 1 660.3 720.3  
## - Hyperlipidemia 1 661.0 721.0  
## - BackPain 1 661.0 721.0  
## - Overweight 1 661.1 721.1  
## - Soft\_drink 1 661.1 721.1  
## <none> 659.4 721.4  
## - Allergic\_rhinitis 1 661.7 721.7  
## - Children 1 662.9 722.9  
## - Reflux\_esophagitis 1 663.2 723.2  
## - Diabetes 1 663.4 723.4  
## - Anxiety 1 682.3 742.3  
## - Asthma 1 697.1 757.1  
## - Arthritis 1 699.2 759.2  
## - Stroke 1 700.3 760.3  
## - Complication\_risk 2 704.5 762.5  
## - Services 3 708.7 764.7  
## - Initial\_admin 2 765.5 823.5  
## - Initial\_days 1 13112.0 13172.0  
##   
## Step: AIC=718.32  
## ReAdmis ~ Initial\_days + Children + Age + Income + VitD\_levels +   
## Doc\_visits + Full\_meals\_eaten + vitD\_supp + Soft\_drink +   
## Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Doc\_visits 1 660.3 716.3  
## - Income 1 660.4 716.4  
## - VitD\_levels 1 660.5 716.5  
## - Full\_meals\_eaten 1 660.6 716.6  
## - Age 1 660.6 716.6  
## - vitD\_supp 1 660.7 716.7  
## - Additional\_charges 1 660.8 716.8  
## - HighBlood 1 661.2 717.2  
## - Hyperlipidemia 1 661.8 717.8  
## - Overweight 1 662.0 718.0  
## - Soft\_drink 1 662.0 718.0  
## - BackPain 1 662.0 718.0  
## <none> 660.3 718.3  
## - Allergic\_rhinitis 1 662.7 718.7  
## - Diabetes 1 664.2 720.2  
## - Children 1 664.2 720.2  
## - Reflux\_esophagitis 1 664.4 720.4  
## - Anxiety 1 683.8 739.8  
## - Asthma 1 697.7 753.7  
## - Arthritis 1 699.5 755.5  
## - Stroke 1 702.4 758.4  
## - Complication\_risk 2 706.6 760.6  
## - Services 3 709.6 761.6  
## - Initial\_admin 2 765.7 819.7  
## - Initial\_days 1 13113.8 13169.8  
##   
## Step: AIC=716.33  
## ReAdmis ~ Initial\_days + Children + Age + Income + VitD\_levels +   
## Full\_meals\_eaten + vitD\_supp + Soft\_drink + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Overweight + Arthritis +   
## Diabetes + Hyperlipidemia + BackPain + Anxiety + Allergic\_rhinitis +   
## Reflux\_esophagitis + Asthma + Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Income 1 660.4 714.4  
## - VitD\_levels 1 660.5 714.5  
## - Full\_meals\_eaten 1 660.6 714.6  
## - Age 1 660.6 714.6  
## - vitD\_supp 1 660.7 714.7  
## - Additional\_charges 1 660.8 714.8  
## - HighBlood 1 661.2 715.2  
## - Hyperlipidemia 1 661.8 715.8  
## - Overweight 1 662.0 716.0  
## - Soft\_drink 1 662.0 716.0  
## - BackPain 1 662.1 716.1  
## <none> 660.3 716.3  
## - Allergic\_rhinitis 1 662.7 716.7  
## - Children 1 664.3 718.3  
## - Diabetes 1 664.3 718.3  
## - Reflux\_esophagitis 1 664.4 718.4  
## - Anxiety 1 683.9 737.9  
## - Asthma 1 697.7 751.7  
## - Arthritis 1 699.5 753.5  
## - Stroke 1 702.4 756.4  
## - Complication\_risk 2 706.6 758.6  
## - Services 3 709.6 759.6  
## - Initial\_admin 2 766.1 818.1  
## - Initial\_days 1 13113.8 13167.8  
##   
## Step: AIC=714.43  
## ReAdmis ~ Initial\_days + Children + Age + VitD\_levels + Full\_meals\_eaten +   
## vitD\_supp + Soft\_drink + Initial\_admin + HighBlood + Stroke +   
## Complication\_risk + Overweight + Arthritis + Diabetes + Hyperlipidemia +   
## BackPain + Anxiety + Allergic\_rhinitis + Reflux\_esophagitis +   
## Asthma + Services + Additional\_charges  
##   
## Df Deviance AIC  
## - VitD\_levels 1 660.6 712.6  
## - Full\_meals\_eaten 1 660.7 712.7  
## - Age 1 660.7 712.7  
## - vitD\_supp 1 660.8 712.8  
## - Additional\_charges 1 660.9 712.9  
## - HighBlood 1 661.3 713.3  
## - Hyperlipidemia 1 662.0 714.0  
## - Overweight 1 662.1 714.1  
## - Soft\_drink 1 662.2 714.2  
## - BackPain 1 662.2 714.2  
## <none> 660.4 714.4  
## - Allergic\_rhinitis 1 662.8 714.8  
## - Diabetes 1 664.3 716.3  
## - Children 1 664.4 716.4  
## - Reflux\_esophagitis 1 664.4 716.4  
## - Anxiety 1 684.0 736.0  
## - Asthma 1 697.7 749.7  
## - Arthritis 1 699.6 751.6  
## - Stroke 1 702.4 754.4  
## - Complication\_risk 2 706.6 756.6  
## - Services 3 709.9 757.9  
## - Initial\_admin 2 766.1 816.1  
## - Initial\_days 1 13115.0 13167.0  
##   
## Step: AIC=712.61  
## ReAdmis ~ Initial\_days + Children + Age + Full\_meals\_eaten +   
## vitD\_supp + Soft\_drink + Initial\_admin + HighBlood + Stroke +   
## Complication\_risk + Overweight + Arthritis + Diabetes + Hyperlipidemia +   
## BackPain + Anxiety + Allergic\_rhinitis + Reflux\_esophagitis +   
## Asthma + Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Full\_meals\_eaten 1 660.9 710.9  
## - Age 1 660.9 710.9  
## - vitD\_supp 1 661.0 711.0  
## - Additional\_charges 1 661.1 711.1  
## - HighBlood 1 661.5 711.5  
## - Hyperlipidemia 1 662.2 712.2  
## - Overweight 1 662.3 712.3  
## - BackPain 1 662.4 712.4  
## - Soft\_drink 1 662.4 712.4  
## <none> 660.6 712.6  
## - Allergic\_rhinitis 1 663.1 713.1  
## - Diabetes 1 664.4 714.4  
## - Reflux\_esophagitis 1 664.6 714.6  
## - Children 1 664.8 714.8  
## - Anxiety 1 684.1 734.1  
## - Asthma 1 697.9 747.9  
## - Arthritis 1 699.6 749.6  
## - Stroke 1 702.7 752.7  
## - Complication\_risk 2 707.6 755.6  
## - Services 3 710.2 756.2  
## - Initial\_admin 2 767.0 815.0  
## - Initial\_days 1 13115.2 13165.2  
##   
## Step: AIC=710.86  
## ReAdmis ~ Initial\_days + Children + Age + vitD\_supp + Soft\_drink +   
## Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Age 1 661.2 709.2  
## - vitD\_supp 1 661.3 709.3  
## - Additional\_charges 1 661.3 709.3  
## - HighBlood 1 661.7 709.7  
## - Hyperlipidemia 1 662.4 710.4  
## - Overweight 1 662.5 710.5  
## - BackPain 1 662.6 710.6  
## - Soft\_drink 1 662.7 710.7  
## <none> 660.9 710.9  
## - Allergic\_rhinitis 1 663.3 711.3  
## - Diabetes 1 664.7 712.7  
## - Reflux\_esophagitis 1 664.8 712.8  
## - Children 1 665.1 713.1  
## - Anxiety 1 684.2 732.2  
## - Asthma 1 698.0 746.0  
## - Arthritis 1 699.7 747.7  
## - Stroke 1 703.0 751.0  
## - Complication\_risk 2 707.9 753.9  
## - Services 3 710.5 754.5  
## - Initial\_admin 2 767.5 813.5  
## - Initial\_days 1 13116.5 13164.5  
##   
## Step: AIC=709.16  
## ReAdmis ~ Initial\_days + Children + vitD\_supp + Soft\_drink +   
## Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services + Additional\_charges  
##   
## Df Deviance AIC  
## - Additional\_charges 1 661.4 707.4  
## - vitD\_supp 1 661.6 707.6  
## - Hyperlipidemia 1 662.8 708.8  
## - Overweight 1 662.8 708.8  
## - BackPain 1 663.0 709.0  
## - Soft\_drink 1 663.1 709.1  
## <none> 661.2 709.2  
## - Allergic\_rhinitis 1 663.6 709.6  
## - Reflux\_esophagitis 1 665.0 711.0  
## - Diabetes 1 665.1 711.1  
## - Children 1 665.4 711.4  
## - HighBlood 1 669.3 715.3  
## - Anxiety 1 684.9 730.9  
## - Asthma 1 698.3 744.3  
## - Arthritis 1 700.4 746.4  
## - Stroke 1 705.2 751.2  
## - Services 3 711.1 753.1  
## - Complication\_risk 2 709.8 753.8  
## - Initial\_admin 2 773.8 817.8  
## - Initial\_days 1 13116.7 13162.7  
##   
## Step: AIC=707.41  
## ReAdmis ~ Initial\_days + Children + vitD\_supp + Soft\_drink +   
## Initial\_admin + HighBlood + Stroke + Complication\_risk +   
## Overweight + Arthritis + Diabetes + Hyperlipidemia + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services  
##   
## Df Deviance AIC  
## - vitD\_supp 1 661.8 705.8  
## - Hyperlipidemia 1 663.0 707.0  
## - Overweight 1 663.1 707.1  
## - BackPain 1 663.2 707.2  
## - Soft\_drink 1 663.3 707.3  
## <none> 661.4 707.4  
## - Allergic\_rhinitis 1 663.9 707.9  
## - Diabetes 1 665.2 709.2  
## - Reflux\_esophagitis 1 665.5 709.5  
## - Children 1 665.6 709.6  
## - HighBlood 1 679.0 723.0  
## - Anxiety 1 685.0 729.0  
## - Asthma 1 698.5 742.5  
## - Arthritis 1 700.5 744.5  
## - Stroke 1 705.8 749.8  
## - Services 3 711.4 751.4  
## - Complication\_risk 2 710.1 752.1  
## - Initial\_admin 2 774.8 816.8  
## - Initial\_days 1 13118.9 13162.9  
##   
## Step: AIC=705.78  
## ReAdmis ~ Initial\_days + Children + Soft\_drink + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Overweight + Arthritis +   
## Diabetes + Hyperlipidemia + BackPain + Anxiety + Allergic\_rhinitis +   
## Reflux\_esophagitis + Asthma + Services  
##   
## Df Deviance AIC  
## - Hyperlipidemia 1 663.3 705.3  
## - Overweight 1 663.4 705.4  
## - BackPain 1 663.6 705.6  
## - Soft\_drink 1 663.8 705.8  
## <none> 661.8 705.8  
## - Allergic\_rhinitis 1 664.3 706.3  
## - Diabetes 1 665.6 707.6  
## - Reflux\_esophagitis 1 665.9 707.9  
## - Children 1 666.2 708.2  
## - HighBlood 1 679.5 721.5  
## - Anxiety 1 685.1 727.1  
## - Asthma 1 698.9 740.9  
## - Arthritis 1 701.2 743.2  
## - Stroke 1 705.8 747.8  
## - Services 3 711.7 749.7  
## - Complication\_risk 2 710.1 750.1  
## - Initial\_admin 2 775.5 815.5  
## - Initial\_days 1 13120.1 13162.1  
##   
## Step: AIC=705.3  
## ReAdmis ~ Initial\_days + Children + Soft\_drink + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Overweight + Arthritis +   
## Diabetes + BackPain + Anxiety + Allergic\_rhinitis + Reflux\_esophagitis +   
## Asthma + Services  
##   
## Df Deviance AIC  
## - Overweight 1 664.8 704.8  
## - BackPain 1 665.1 705.1  
## <none> 663.3 705.3  
## - Soft\_drink 1 665.3 705.3  
## - Allergic\_rhinitis 1 666.0 706.0  
## - Diabetes 1 667.1 707.1  
## - Reflux\_esophagitis 1 667.6 707.6  
## - Children 1 667.6 707.6  
## - HighBlood 1 681.0 721.0  
## - Anxiety 1 687.0 727.0  
## - Asthma 1 700.6 740.6  
## - Arthritis 1 702.2 742.2  
## - Stroke 1 706.5 746.5  
## - Services 3 713.8 749.8  
## - Complication\_risk 2 713.0 751.0  
## - Initial\_admin 2 778.2 816.2  
## - Initial\_days 1 13120.2 13160.2  
##   
## Step: AIC=704.81  
## ReAdmis ~ Initial\_days + Children + Soft\_drink + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Arthritis + Diabetes +   
## BackPain + Anxiety + Allergic\_rhinitis + Reflux\_esophagitis +   
## Asthma + Services  
##   
## Df Deviance AIC  
## - Soft\_drink 1 666.6 704.6  
## - BackPain 1 666.8 704.8  
## <none> 664.8 704.8  
## - Allergic\_rhinitis 1 667.4 705.4  
## - Diabetes 1 668.7 706.7  
## - Reflux\_esophagitis 1 668.7 706.7  
## - Children 1 669.3 707.3  
## - HighBlood 1 681.7 719.7  
## - Anxiety 1 688.6 726.6  
## - Asthma 1 701.9 739.9  
## - Arthritis 1 704.6 742.6  
## - Stroke 1 707.9 745.9  
## - Services 3 715.0 749.0  
## - Complication\_risk 2 713.9 749.9  
## - Initial\_admin 2 779.1 815.1  
## - Initial\_days 1 13120.9 13158.9  
##   
## Step: AIC=704.6  
## ReAdmis ~ Initial\_days + Children + Initial\_admin + HighBlood +   
## Stroke + Complication\_risk + Arthritis + Diabetes + BackPain +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services  
##   
## Df Deviance AIC  
## - BackPain 1 668.5 704.5  
## <none> 666.6 704.6  
## - Allergic\_rhinitis 1 669.4 705.4  
## - Reflux\_esophagitis 1 670.6 706.6  
## - Diabetes 1 670.7 706.7  
## - Children 1 671.1 707.1  
## - HighBlood 1 683.5 719.5  
## - Anxiety 1 689.8 725.8  
## - Asthma 1 703.0 739.0  
## - Arthritis 1 705.9 741.9  
## - Stroke 1 709.4 745.4  
## - Complication\_risk 2 714.3 748.3  
## - Services 3 716.8 748.8  
## - Initial\_admin 2 779.7 813.7  
## - Initial\_days 1 13121.4 13157.4  
##   
## Step: AIC=704.51  
## ReAdmis ~ Initial\_days + Children + Initial\_admin + HighBlood +   
## Stroke + Complication\_risk + Arthritis + Diabetes + Anxiety +   
## Allergic\_rhinitis + Reflux\_esophagitis + Asthma + Services  
##   
## Df Deviance AIC  
## <none> 668.5 704.5  
## - Allergic\_rhinitis 1 671.2 705.2  
## - Reflux\_esophagitis 1 672.3 706.3  
## - Diabetes 1 672.6 706.6  
## - Children 1 673.0 707.0  
## - HighBlood 1 685.8 719.8  
## - Anxiety 1 690.8 724.8  
## - Asthma 1 704.7 738.7  
## - Arthritis 1 708.4 742.4  
## - Stroke 1 710.7 744.7  
## - Complication\_risk 2 716.5 748.5  
## - Services 3 719.5 749.5  
## - Initial\_admin 2 782.4 814.4  
## - Initial\_days 1 13123.1 13157.1

# Show forward-reduced model summary  
summary(stepfwd)

##   
## Call:  
## glm(formula = ReAdmis ~ Initial\_days + Initial\_admin + Services +   
## Stroke + Complication\_risk + Arthritis + Asthma + Anxiety +   
## HighBlood + Diabetes + Children + Reflux\_esophagitis + Allergic\_rhinitis,   
## family = binomial, data = init\_mdl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.5137 0.0000 0.0000 0.0001 3.3847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -80.77379 4.51401 -17.894 < 2e-16 \*\*\*  
## Initial\_days 1.47447 0.08229 17.918 < 2e-16 \*\*\*  
## Initial\_adminEmergency Admission 2.52225 0.27709 9.103 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 0.77122 0.27467 2.808 0.00499 \*\*   
## ServicesCT Scan 1.58311 0.35925 4.407 1.05e-05 \*\*\*  
## ServicesIntravenous 0.01598 0.22706 0.070 0.94391   
## ServicesMRI 2.74402 0.48841 5.618 1.93e-08 \*\*\*  
## StrokeYes 1.62981 0.26347 6.186 6.17e-10 \*\*\*  
## Complication\_riskLow -1.85203 0.29084 -6.368 1.92e-10 \*\*\*  
## Complication\_riskMedium -0.33849 0.22983 -1.473 0.14081   
## ArthritisYes -1.34063 0.22254 -6.024 1.70e-09 \*\*\*  
## AsthmaYes -1.34759 0.23301 -5.783 7.32e-09 \*\*\*  
## AnxietyYes -1.02489 0.22300 -4.596 4.31e-06 \*\*\*  
## HighBloodYes 0.86692 0.21263 4.077 4.56e-05 \*\*\*  
## DiabetesYes 0.46011 0.22806 2.018 0.04364 \*   
## Children 0.09512 0.04528 2.101 0.03566 \*   
## Reflux\_esophagitisYes -0.40251 0.20678 -1.947 0.05158 .   
## Allergic\_rhinitisYes -0.33276 0.20464 -1.626 0.10394   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 13145.70 on 9999 degrees of freedom  
## Residual deviance: 668.51 on 9982 degrees of freedom  
## AIC: 704.51  
##   
## Number of Fisher Scoring iterations: 12

# Show backward-reduced model summary  
summary(stepbck)

##   
## Call:  
## glm(formula = ReAdmis ~ Initial\_days + Children + Initial\_admin +   
## HighBlood + Stroke + Complication\_risk + Arthritis + Diabetes +   
## Anxiety + Allergic\_rhinitis + Reflux\_esophagitis + Asthma +   
## Services, family = binomial, data = init\_mdl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.5137 0.0000 0.0000 0.0001 3.3847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -80.77379 4.51401 -17.894 < 2e-16 \*\*\*  
## Initial\_days 1.47447 0.08229 17.918 < 2e-16 \*\*\*  
## Children 0.09512 0.04528 2.101 0.03566 \*   
## Initial\_adminEmergency Admission 2.52225 0.27709 9.103 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 0.77122 0.27467 2.808 0.00499 \*\*   
## HighBloodYes 0.86692 0.21263 4.077 4.56e-05 \*\*\*  
## StrokeYes 1.62981 0.26347 6.186 6.17e-10 \*\*\*  
## Complication\_riskLow -1.85203 0.29084 -6.368 1.92e-10 \*\*\*  
## Complication\_riskMedium -0.33849 0.22983 -1.473 0.14081   
## ArthritisYes -1.34063 0.22254 -6.024 1.70e-09 \*\*\*  
## DiabetesYes 0.46011 0.22806 2.018 0.04364 \*   
## AnxietyYes -1.02489 0.22300 -4.596 4.31e-06 \*\*\*  
## Allergic\_rhinitisYes -0.33276 0.20464 -1.626 0.10394   
## Reflux\_esophagitisYes -0.40251 0.20678 -1.947 0.05158 .   
## AsthmaYes -1.34759 0.23301 -5.783 7.32e-09 \*\*\*  
## ServicesCT Scan 1.58311 0.35925 4.407 1.05e-05 \*\*\*  
## ServicesIntravenous 0.01598 0.22706 0.070 0.94391   
## ServicesMRI 2.74402 0.48841 5.618 1.93e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 13145.70 on 9999 degrees of freedom  
## Residual deviance: 668.51 on 9982 degrees of freedom  
## AIC: 704.51  
##   
## Number of Fisher Scoring iterations: 12

# Reduced Model (stepbck)  
reduced\_mdl <- glm(ReAdmis ~ Initial\_days +  
 Initial\_admin +  
 Services +  
 Stroke +  
 Complication\_risk +  
 Arthritis +  
 Asthma +  
 Anxiety +  
 HighBlood +  
 Diabetes +  
 Children +  
 Reflux\_esophagitis +  
 Allergic\_rhinitis,  
 data = init\_mdl,  
 family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Part VI

### F1: Results

Multiple logistic regression analyses were performed to interrogate the relationship between the binary response variable, ReAdmis, and a group of various, potentially related predictor variables as the initial model. Subsequently, two variable selection methods were employed, namely forward selection and backward elimination, to iteratively reduce the initial model. The regression equation for the reduced model is:

As mentioned above, the logistic regression equation is as follows: (Ciaburro, 2018)

Therefore, a potential interpretation of this model and its coefficients could be:

However, in order to derive real meaning from the model, it’s useful to convert the coefficients of the model from log-odds to probability. That conversion looks like the following:

Statistically, this model is quite reliable in predicting patient readmission based on the given parameters. The confusion matrix above details a highly accurate classification in terms of true positive and negative predictions. Practically, however, the model relies heavily on Initial\_days as a predictor. This seems to indicate the longer a patient has been initially hospitalized, the higher the likelihood of their eventual readmission. This cannot be a surprise to those familiar with healthcare outcomes. Therefore, the value of the model is somewhat diminished, in my opinion, due to how heavily it relies on that single feature. Nonetheless, the model on the whole is not overly burdensome and it is quite accurate, at least with regard to the dataset at hand.

Additionally, it should be noted that the Initial\_days data is less than ideal given that its distribution is not normal. If anything, the distribution seems to take a bimodal shape and seems to have an odd cutoff point with no major outliers on the high end. It’s possible the skew as well as the shape of the data caused the model to produce a “perfect separation” warning when fitting any combination of features including Initial\_days or simply ReAdmis ~ Initial\_days itself.

### F2: Recommendations

As a result of the preceding analysis, it is recommended that leadership consider implementing policy and procedure utilizing this logistic regression model to estimate a patient’s potential probability of readmission. This model can be used as a limited, additional tool to estimate, and attempt to mitigate, readmission risk which would likely lead to better outcomes overall.

Furthermore, it is recommended that additional, prospective analysis and data capture be performed to better understand the distribution, as well as what potential drivers could be behind the distribution, of the Initial\_days data. Also, other continuous variables of interest in future data capture could include various lab values such as those found in blood gas analysis, complete blood count panel, comprehensive metabolic panel, etc. As these laboratory values are obtained frequently and during hospitalization, and in some cases during prior doctor visits, and are highly indicative of a patient’s status, it seems likely that having access to these data would allow for more useful, predictive models.

# I: Sources

Anderson, D. (2002). *Model selection and multimodel inference* (2 ed.). New York: Springer.

Ciaburro, G. (2018). *Regression Analysis with R.* Birmingham: Packt.