D208 - Predictive Modeling

Table of Contents

#### **Performance Assessment - Task 1: Multiple Regression for Predictive Modeling**

#### *Medical Readmission Data Set (Clean)*

## 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 initial total days spent hospitalized (Initial\_days)?

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

No combination of variables included in the medical dataset influence initial length of stay (Initial\_days) in any statistically significant way.

Additionally, our alternate hypothesis () is:

Some combination of variables included in the medical dataset influence initial length of stay (Initial\_days) in a statistically significant way.

### A2: Objectives and Goals

The primary goal of the following analysis is to determine what variables, if any, contribute positively to a patient’s initial number of days hospitalized (Initial\_days). This will be assessed using the programming language and using the technique of multiple linear regression to identify causal relationships between independent variables and the target variable.

## Part II

### B1: Summary of Assumptions

The following are basic assumptions of multiple linear regression:

* The assumption of a linear relationship between the explanatory variables and the target variable
* The assumption that residual values are normally distributed
* The assumption of non-multicollinearity of explanatory variables
* The assumption that residual values are homoscedastic

### 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 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 ols\_step\_backward\_p() 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 regression is an appropriate technique to use to accomplish our goal of finding which variables contribute to a longer initial length of stay (Initial\_days) for several reasons. Firstly, the dataset we will analyze contains 50 variables in total, each of disparate significance and utility. While some variable do not require much thought prior to elimination, others are not quite as straightforward. Therefore, running them through a multiple 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 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 take 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 model selection process we will use is backward-oriented, this initial model will include all features that could possibly have a relationship to the target variable of initial length of hospitalization (Initial\_days). 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 categorical variables are nominal and which would benefit from ordinal encoding. 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: |  |
| 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 |

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))

# Read dataset in  
df <- read.csv("./data/medical\_clean.csv")  
  
# 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 %>%  
 select(Initial\_days,  
 Area,  
 Children:Services,  
 Additional\_charges)  
  
# View new dataframe and assess dtypes  
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 dataframes 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. We do, however, have a couple of variables that likely would benefit from changing the reference category. The previous code was convenient for converting a relatively large group of variables to factors all at once, but does not consider levels. Thus, we will take to changing the reference level for two variables (Initial\_admin and Complication\_risk) below. First, let’s just verify our assumptions by accessing these variables’ levels to determine which category comes first. If our model is attempting to discover what features contribute to a longer initial length of hospitalization, it would be helpful to have the least likely contributor be the reference category, such that the coefficients are positively oriented. Based on measures of central tendency for these two variables, Emergency Admission for Initial\_admin and Medium Complication\_risk have the lowest combination of median, mean and standard deviation and will be reset as the reference category for their respective variable.

# Access levels() for each variable  
levels(init\_mdl$Initial\_admin)

## [1] "Elective Admission" "Emergency Admission" "Observation Admission"

levels(init\_mdl$Complication\_risk)

## [1] "High" "Low" "Medium"

OK, our assumptions about the order were valid. Now, let’s fix them.

# Re-level ordinal categorical variables for Initial\_admin  
init\_mdl$Initial\_admin <- factor(init\_mdl$Initial\_admin,  
 levels = c("Emergency Admission",  
 "Observation Admission",  
 "Elective Admission"))  
# Re-level ordinal categorical variables for Complication\_risk  
init\_mdl$Complication\_risk <- factor(init\_mdl$Complication\_risk,  
 levels = c("Medium",  
 "Low",  
 "High"))

Finally, we will check again to make sure the above worked as intended.

# Access levels() for each variable  
levels(init\_mdl$Initial\_admin)

## [1] "Emergency Admission" "Observation Admission" "Elective Admission"

levels(init\_mdl$Complication\_risk)

## [1] "Medium" "Low" "High"

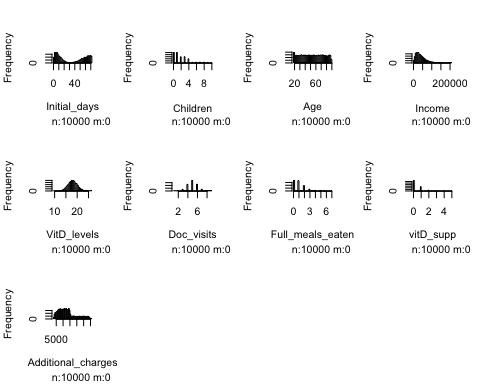
Now that we have validated our conversion process and the composition of our data, we can now proceed with the model selection process.

### C4: Visualizations

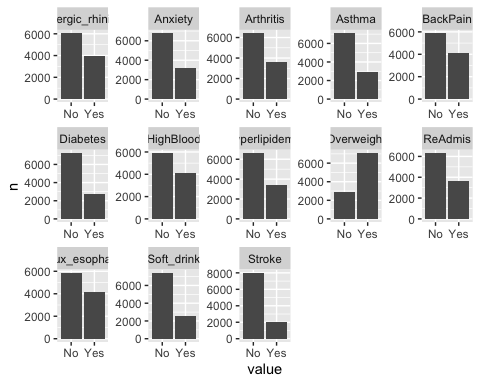
#### Univariate:

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))

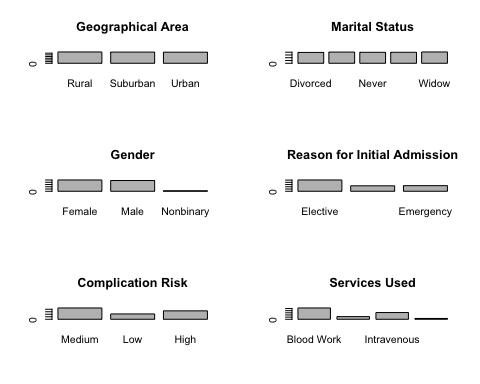


# Partition dichotomous Yes/No variables out for plot  
dichotomous\_vars <- init\_mdl %>%   
 select(where(~n\_distinct(.) == 2))  
  
# Show bar charts of all Yes/No variables  
dichotomous\_vars %>%  
 gather() %>%  
 count(key, value) %>%   
 ggplot(., aes(x = value, y = n)) +  
 geom\_bar(stat =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



# Partition non-dichotomous categorical variables  
cat\_vars <- init\_mdl %>%   
 select\_if(is.factor) %>%   
 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")

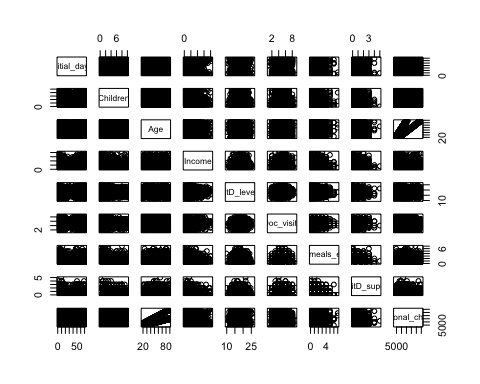
# Create panel of bar charts for cat\_vars  
par(mfrow = c(3,2))  
barplot(table(cat\_vars$Area), main = "Geographical Area")  
barplot(table(cat\_vars$Marital), main = "Marital Status")  
barplot(table(cat\_vars$Gender), main = "Gender")  
barplot(table(cat\_vars$Initial\_admin), main = "Reason for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



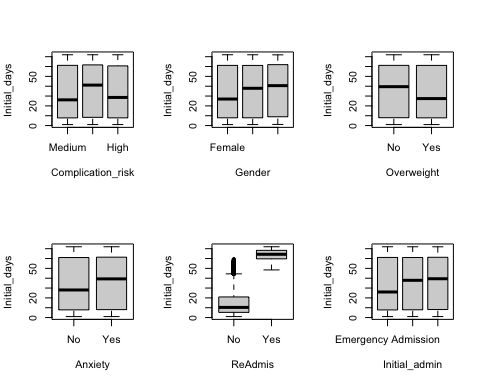
#### Bivariate:

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 scatterplot matrix of numeric variables  
pairs(num\_vars)



# Boxplots of our target variable against some categorical variables  
par(mfrow = c(2,3))  
boxplot(Initial\_days ~ Complication\_risk, data = init\_mdl)  
boxplot(Initial\_days ~ Gender, data = init\_mdl)  
boxplot(Initial\_days ~ Overweight, data = init\_mdl)  
boxplot(Initial\_days ~ Anxiety, data = init\_mdl)  
boxplot(Initial\_days ~ ReAdmis, data = init\_mdl)  
boxplot(Initial\_days ~ Initial\_admin, data = init\_mdl)



### 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 a combination of backward and stepwise selection. Considering the linear 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 regression formula in , we’ve already cleaned and transformed our dataframe for the initial model, so the formula is much more concise:

Initial\_days ~ .,  
 data = init\_mdl)

### D2: Justification of Model Reduction

Our variable selection process includes the use of three stepwise elimination methods, namely backward, forward, and bidirectional stepwise elimination. The three selection methods are similar in approach, but each can result in substantively different models. Using all three will, 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 t-statistic p-value and whether or not it meets the threshold of <0.1 for a given feature to be selected. Then, we will compare and evaluate the three resulting models using their corresponding Root Mean Square Error (RMSE), , and F-statistic. The model with the “best” combination of these measures (i.e. lowest RMSE, highest , and lowest F-statistic) will be selected as our reduced model.

### D3: Reduced Multiple Regression Model

The reduced model is as follows:

Our multiple regression formula in , is:

Initial\_days ~ Age +   
 ReAdmis +   
 Initial\_admin +   
 Complication\_risk +   
 Arthritis +   
 Anxiety +   
 Additional\_charges,   
 data = init\_mdl)

## Part V

### E1: Model Comparison

As mentioned above, the process by which the multiple regression model was reduced and selected was using backward, forward, and bidirectional stepwise selection. Using the olsrr package, we were able to run the selection process as an iterative function and analyze the results. Additionally, each step in the process is tracked and all decisions made by the algorithm are recorded. The initial model was a completely overfit model with very little useful information to begin with. The backward elimination process analyzed the initial model, compared the t-statistic p-values for all features, and removed the feature with the least statistically significant contribution to the model at each step. This resulted in 20 total steps eliminating variables until no variable in the model had a p-value >0.1.

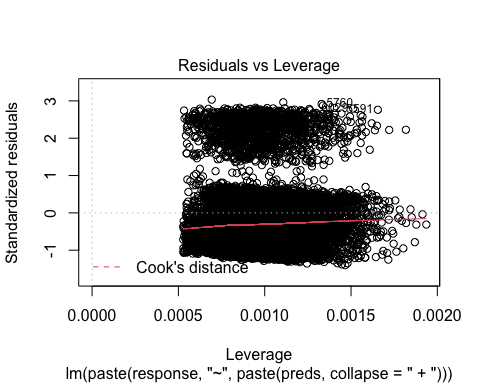
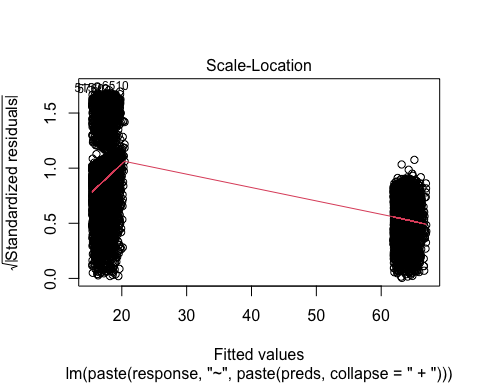
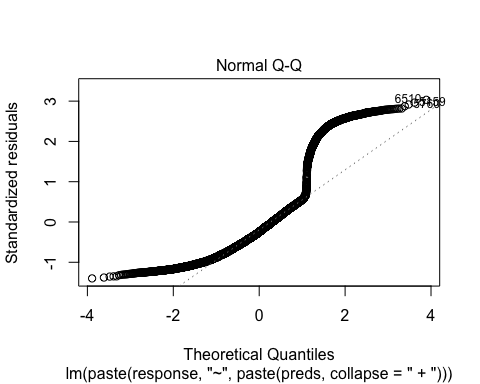
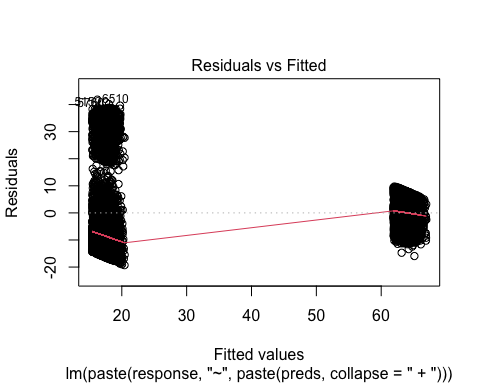
The forward selection process used the same evaluation criterion (p-value <0.1). The process was simply the reverse of backward elimination. Analysis of the initial model’s features resulted in the identification of the most statistically significant variable, adding it to the model. This step was repeated, each time adding the next-most statistically significant variable until none of the remaining variables met the p-value threshold of <0.1.

Finally, the bidirectional method was a combination of both forward and backward stepwise procedures. The algorithm starts with forward selection and at each step evaluates the most significant impact to the model in whether or not to select or eliminate a variable until the criteria (again, p-value <0.1) is met.

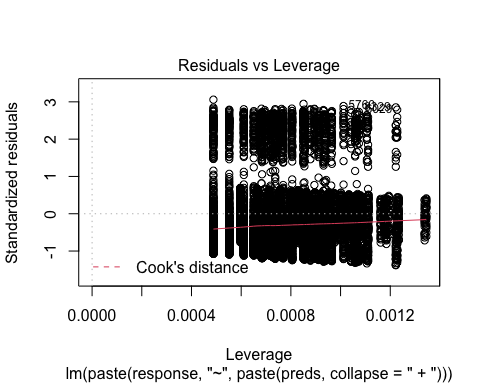
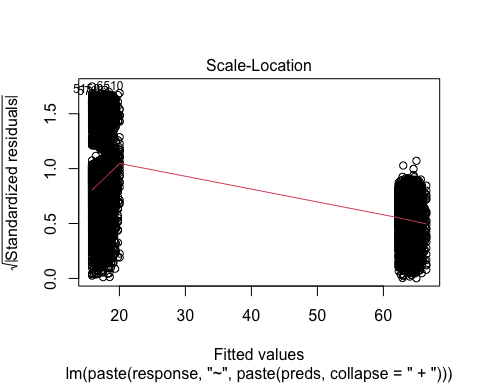
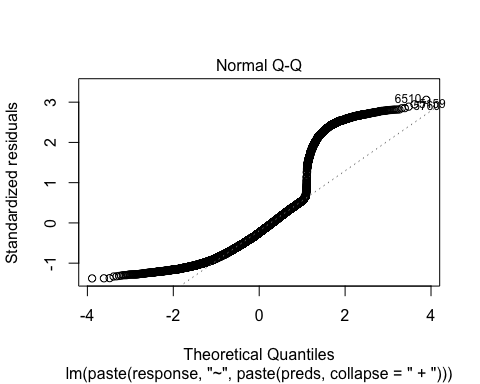
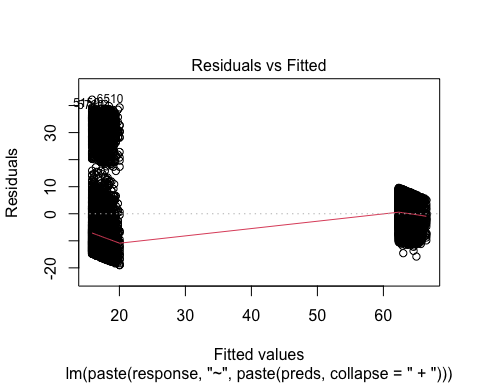
The final output of the three models includes the RMSE, adjusted , and F-statistic. The models were then compared by these metrics. There is very little different between models. In fact, the forward and bidirectional methods resulted in the same model, so the evaluation was essentially between the backward and forward methods. As the backward elimination method produced a model with a lower F-statistic and the RMSE and adjusted were very similar, the backward elimination model was chosen as the final reduced model.

Consider the following sets of plots for both models:

# Plot residual analysis of backward model  
plot(mdl\_back$model)



# Plot residual analysis of forward model  
plot(mdl\_fwd$model)



### E2: Output and Calculations

Consider the following summary outputs for both models including the residual standard error, adjusted , and F-statistic.

# Summary of the backward model  
summary(mdl\_back$model)

##   
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),   
## data = l)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.296 -9.671 -3.406 4.212 41.812   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.768e+01 5.193e-01 34.050 < 2e-16 \*\*\*  
## Age 1.760e-02 9.603e-03 1.833 0.0668 .   
## ReAdmisYes 4.647e+01 2.862e-01 162.344 < 2e-16 \*\*\*  
## Initial\_adminEmergency Admission -1.590e+00 3.374e-01 -4.712 2.48e-06 \*\*\*  
## Initial\_adminObservation Admission -2.585e-01 3.926e-01 -0.658 0.5103   
## Complication\_riskLow 9.479e-01 3.831e-01 2.474 0.0134 \*   
## Complication\_riskMedium -3.323e-01 3.146e-01 -1.056 0.2909   
## ArthritisYes 6.747e-01 2.879e-01 2.344 0.0191 \*   
## AnxietyYes 5.699e-01 2.953e-01 1.930 0.0536 .   
## Additional\_charges -6.383e-05 3.034e-05 -2.104 0.0354 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9990 degrees of freedom  
## Multiple R-squared: 0.7255, Adjusted R-squared: 0.7253   
## F-statistic: 2934 on 9 and 9990 DF, p-value: < 2.2e-16

# Summary of the forward model  
summary(mdl\_fwd$model)

##   
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),   
## data = l)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.000 -9.647 -3.375 4.218 42.178   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.7952 0.3744 47.526 < 2e-16 \*\*\*  
## ReAdmisYes 46.4686 0.2862 162.346 < 2e-16 \*\*\*  
## Initial\_adminEmergency Admission -1.6229 0.3371 -4.815 1.5e-06 \*\*\*  
## Initial\_adminObservation Admission -0.2598 0.3926 -0.662 0.50821   
## Complication\_riskLow 0.9982 0.3824 2.610 0.00906 \*\*   
## Complication\_riskMedium -0.3096 0.3145 -0.985 0.32484   
## ArthritisYes 0.6754 0.2879 2.346 0.01899 \*   
## AnxietyYes 0.5645 0.2953 1.912 0.05596 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9992 degrees of freedom  
## Multiple R-squared: 0.7254, Adjusted R-squared: 0.7252   
## F-statistic: 3770 on 7 and 9992 DF, p-value: < 2.2e-16

### Predictions using reduced model:

# Create a dataframe of 20 predictor vectors  
p\_Age <- seq.int(18, 95, by = 4)  
p\_ReAdmis <- rep\_len(levels(init\_mdl$ReAdmis), length.out = 20)  
p\_Initial\_admin <- rep\_len(levels(init\_mdl$Initial\_admin), 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\_Anxiety <- rep\_len(levels(init\_mdl$Anxiety), length.out = 20)  
p\_Additional\_charges <- seq(min(init\_mdl$Additional\_charges), max(init\_mdl$Additional\_charges), length.out = 20)  
p\_mdl <- data.frame(Age = p\_Age,  
 ReAdmis = p\_ReAdmis,  
 Initial\_admin = p\_Initial\_admin,  
 Complication\_risk = p\_Complication\_risk,  
 Arthritis = p\_Arthritis,  
 Anxiety = p\_Anxiety,  
 Additional\_charges = p\_Additional\_charges)  
p\_mdl$p\_initial\_days <- predict(mdl\_back$model, new = p\_mdl)  
p\_mdl

## Age ReAdmis Initial\_admin Complication\_risk Arthritis Anxiety  
## 1 18 No Emergency Admission Medium No No  
## 2 22 Yes Observation Admission Low Yes Yes  
## 3 26 No Elective Admission High No No  
## 4 30 Yes Emergency Admission Medium Yes Yes  
## 5 34 No Observation Admission Low No No  
## 6 38 Yes Elective Admission High Yes Yes  
## 7 42 No Emergency Admission Medium No No  
## 8 46 Yes Observation Admission Low Yes Yes  
## 9 50 No Elective Admission High No No  
## 10 54 Yes Emergency Admission Medium Yes Yes  
## 11 58 No Observation Admission Low No No  
## 12 62 Yes Elective Admission High Yes Yes  
## 13 66 No Emergency Admission Medium No No  
## 14 70 Yes Observation Admission Low Yes Yes  
## 15 74 No Elective Admission High No No  
## 16 78 Yes Emergency Admission Medium Yes Yes  
## 17 82 No Observation Admission Low No No  
## 18 86 Yes Elective Admission High Yes Yes  
## 19 90 No Emergency Admission Medium No No  
## 20 94 Yes Observation Admission Low Yes Yes  
## Additional\_charges p\_initial\_days  
## 1 3125.703 15.87681  
## 2 4569.933 66.17918  
## 3 6014.163 17.75535  
## 4 7458.393 63.52409  
## 5 8902.622 18.40123  
## 6 10346.852 65.40263  
## 7 11791.082 15.74615  
## 8 13235.312 66.04851  
## 9 14679.542 17.62468  
## 10 16123.772 63.39343  
## 11 17568.001 18.27057  
## 12 19012.231 65.27196  
## 13 20456.461 15.61548  
## 14 21900.691 65.91784  
## 15 23344.921 17.49401  
## 16 24789.151 63.26276  
## 17 26233.380 18.13990  
## 18 27677.610 65.14129  
## 19 29121.840 15.48481  
## 20 30566.070 65.78718

### E3: Code

# Assign full initial model (using all previously selected vars) to mdl\_fit  
mdl\_fit <- lm(Initial\_days ~ .,  
 data = init\_mdl)  
# View summary of initial model  
summary(mdl\_fit)

##   
## Call:  
## lm(formula = Initial\_days ~ ., data = init\_mdl)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.902 -9.638 -3.378 4.174 41.291   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.745e+01 1.600e+00 10.909 < 2e-16 \*\*\*  
## AreaSuburban 1.531e-01 3.375e-01 0.454 0.650162   
## AreaUrban 3.422e-01 3.384e-01 1.011 0.311983   
## Children 3.910e-02 6.387e-02 0.612 0.540457   
## Age 3.943e-02 2.035e-02 1.938 0.052696 .   
## Income -3.355e-06 4.843e-06 -0.693 0.488424   
## MaritalMarried -3.622e-02 4.379e-01 -0.083 0.934089   
## MaritalNever Married 4.353e-01 4.400e-01 0.989 0.322510   
## MaritalSeparated 7.838e-01 4.395e-01 1.784 0.074535 .   
## MaritalWidowed 2.832e-01 4.367e-01 0.649 0.516655   
## GenderMale -7.211e-02 2.797e-01 -0.258 0.796552   
## GenderNonbinary -2.700e-01 9.642e-01 -0.280 0.779446   
## ReAdmisYes 4.646e+01 2.867e-01 162.074 < 2e-16 \*\*\*  
## VitD\_levels -7.895e-02 6.854e-02 -1.152 0.249408   
## Doc\_visits -1.666e-01 1.321e-01 -1.261 0.207415   
## Full\_meals\_eaten -1.632e-01 1.371e-01 -1.191 0.233789   
## vitD\_supp 2.691e-01 2.198e-01 1.224 0.220883   
## Soft\_drinkYes -1.814e-01 3.160e-01 -0.574 0.565988   
## Initial\_adminObservation Admission 1.266e+00 3.441e-01 3.679 0.000235 \*\*\*  
## Initial\_adminElective Admission 1.514e+00 3.400e-01 4.453 8.55e-06 \*\*\*  
## HighBloodYes 9.659e-01 7.867e-01 1.228 0.219545   
## StrokeYes -1.172e-01 3.469e-01 -0.338 0.735416   
## Complication\_riskLow 1.215e+00 3.634e-01 3.343 0.000831 \*\*\*  
## Complication\_riskHigh 3.585e-01 3.164e-01 1.133 0.257316   
## OverweightYes -1.938e-01 3.042e-01 -0.637 0.524175   
## ArthritisYes 6.634e-01 2.884e-01 2.301 0.021439 \*   
## DiabetesYes 1.353e-02 3.100e-01 0.044 0.965197   
## HyperlipidemiaYes -3.946e-01 2.923e-01 -1.350 0.176984   
## BackPainYes 3.251e-01 2.811e-01 1.157 0.247393   
## AnxietyYes 5.524e-01 2.957e-01 1.868 0.061765 .   
## Allergic\_rhinitisYes 3.976e-01 2.827e-01 1.407 0.159554   
## Reflux\_esophagitisYes 3.993e-01 2.805e-01 1.424 0.154575   
## AsthmaYes 7.467e-02 3.047e-01 0.245 0.806415   
## ServicesCT Scan -1.059e+00 4.380e-01 -2.418 0.015627 \*   
## ServicesIntravenous -9.154e-02 3.117e-01 -0.294 0.769025   
## ServicesMRI -1.702e-01 7.337e-01 -0.232 0.816538   
## Additional\_charges -1.591e-04 8.512e-05 -1.869 0.061592 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9963 degrees of freedom  
## Multiple R-squared: 0.7263, Adjusted R-squared: 0.7253   
## F-statistic: 734.3 on 36 and 9963 DF, p-value: < 2.2e-16

# Fit backward elimination stepwise regression model to mdl\_back   
mdl\_back <- ols\_step\_backward\_p(mdl\_fit, details = TRUE, prem = .1)

## Backward Elimination Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1 . Area   
## 2 . Children   
## 3 . Age   
## 4 . Income   
## 5 . Marital   
## 6 . Gender   
## 7 . ReAdmis   
## 8 . VitD\_levels   
## 9 . Doc\_visits   
## 10 . Full\_meals\_eaten   
## 11 . vitD\_supp   
## 12 . Soft\_drink   
## 13 . Initial\_admin   
## 14 . HighBlood   
## 15 . Stroke   
## 16 . Complication\_risk   
## 17 . Overweight   
## 18 . Arthritis   
## 19 . Diabetes   
## 20 . Hyperlipidemia   
## 21 . BackPain   
## 22 . Anxiety   
## 23 . Allergic\_rhinitis   
## 24 . Reflux\_esophagitis   
## 25 . Asthma   
## 26 . Services   
## 27 . Additional\_charges   
##   
## We are eliminating variables based on p value...  
##   
## - Diabetes   
##   
## Backward Elimination: Step 1   
##   
## Variable Diabetes Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.789   
## R-Squared 0.726 Coef. Var 40.019   
## Adj. R-Squared 0.725 MSE 190.128   
## Pred R-Squared 0.724 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026685.641 35 143619.590 755.383 0.0000   
## Residual 1894436.581 9964 190.128   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.456 1.596 10.937 0.000 14.328 20.585   
## AreaSuburban 0.153 0.337 0.003 0.453 0.650 -0.508 0.814   
## AreaUrban 0.342 0.338 0.006 1.010 0.312 -0.321 1.005   
## Children 0.039 0.064 0.003 0.613 0.540 -0.086 0.164   
## Age 0.039 0.020 0.031 1.937 0.053 0.000 0.079   
## Income 0.000 0.000 -0.004 -0.693 0.488 0.000 0.000   
## MaritalMarried -0.036 0.438 -0.001 -0.083 0.934 -0.895 0.822   
## MaritalNever Married 0.435 0.440 0.007 0.989 0.323 -0.427 1.298   
## MaritalSeparated 0.784 0.439 0.012 1.784 0.074 -0.078 1.645   
## MaritalWidowed 0.283 0.437 0.004 0.648 0.517 -0.573 1.139   
## GenderMale -0.072 0.280 -0.001 -0.258 0.796 -0.620 0.476   
## GenderNonbinary -0.270 0.964 -0.001 -0.280 0.779 -2.160 1.620   
## ReAdmisYes 46.464 0.287 0.851 162.082 0.000 45.902 47.026   
## VitD\_levels -0.079 0.069 -0.006 -1.153 0.249 -0.213 0.055   
## Doc\_visits -0.166 0.132 -0.007 -1.260 0.208 -0.425 0.092   
## Full\_meals\_eaten -0.163 0.137 -0.006 -1.190 0.234 -0.432 0.106   
## vitD\_supp 0.269 0.220 0.006 1.224 0.221 -0.162 0.700   
## Soft\_drinkYes -0.181 0.316 -0.003 -0.573 0.566 -0.800 0.438   
## Initial\_adminObservation Admission 1.266 0.344 0.021 3.680 0.000 0.592 1.941   
## Initial\_adminElective Admission 1.514 0.340 0.025 4.454 0.000 0.848 2.181   
## HighBloodYes 0.965 0.787 0.018 1.227 0.220 -0.576 2.507   
## StrokeYes -0.117 0.347 -0.002 -0.338 0.736 -0.797 0.563   
## Complication\_riskLow 1.215 0.363 0.019 3.344 0.001 0.503 1.927   
## Complication\_riskHigh 0.358 0.316 0.006 1.133 0.257 -0.262 0.979   
## OverweightYes -0.194 0.304 -0.003 -0.637 0.524 -0.790 0.402   
## ArthritisYes 0.663 0.288 0.012 2.301 0.021 0.098 1.229   
## HyperlipidemiaYes -0.394 0.292 -0.007 -1.350 0.177 -0.967 0.178   
## BackPainYes 0.325 0.281 0.006 1.156 0.248 -0.226 0.876   
## AnxietyYes 0.552 0.296 0.010 1.868 0.062 -0.027 1.132   
## Allergic\_rhinitisYes 0.398 0.283 0.007 1.407 0.159 -0.156 0.952   
## Reflux\_esophagitisYes 0.399 0.280 0.007 1.423 0.155 -0.151 0.949   
## AsthmaYes 0.075 0.305 0.001 0.246 0.806 -0.522 0.672   
## ServicesCT Scan -1.059 0.438 -0.013 -2.418 0.016 -1.917 -0.200   
## ServicesIntravenous -0.091 0.312 -0.002 -0.293 0.769 -0.702 0.520   
## ServicesMRI -0.170 0.733 -0.001 -0.231 0.817 -1.607 1.268   
## Additional\_charges 0.000 0.000 -0.040 -1.869 0.062 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Gender   
##   
## Backward Elimination: Step 2   
##   
## Variable Gender Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.787   
## R-Squared 0.726 Coef. Var 40.015   
## Adj. R-Squared 0.725 MSE 190.092   
## Pred R-Squared 0.724 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026661.476 33 152323.075 801.311 0.0000   
## Residual 1894460.746 9966 190.092   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.410 1.587 10.972 0.000 14.300 20.521   
## AreaSuburban 0.153 0.337 0.003 0.454 0.650 -0.508 0.815   
## AreaUrban 0.341 0.338 0.006 1.009 0.313 -0.322 1.004   
## Children 0.039 0.064 0.003 0.615 0.538 -0.086 0.164   
## Age 0.040 0.020 0.031 1.952 0.051 0.000 0.079   
## Income 0.000 0.000 -0.004 -0.695 0.487 0.000 0.000   
## MaritalMarried -0.035 0.438 -0.001 -0.079 0.937 -0.893 0.823   
## MaritalNever Married 0.438 0.440 0.007 0.995 0.320 -0.424 1.300   
## MaritalSeparated 0.786 0.439 0.012 1.789 0.074 -0.075 1.647   
## MaritalWidowed 0.284 0.437 0.004 0.650 0.516 -0.572 1.140   
## ReAdmisYes 46.462 0.287 0.851 162.108 0.000 45.900 47.024   
## VitD\_levels -0.079 0.068 -0.006 -1.153 0.249 -0.213 0.055   
## Doc\_visits -0.166 0.132 -0.007 -1.259 0.208 -0.425 0.093   
## Full\_meals\_eaten -0.164 0.137 -0.006 -1.194 0.233 -0.432 0.105   
## vitD\_supp 0.268 0.220 0.006 1.219 0.223 -0.163 0.699   
## Soft\_drinkYes -0.181 0.316 -0.003 -0.574 0.566 -0.800 0.438   
## Initial\_adminObservation Admission 1.266 0.344 0.021 3.680 0.000 0.592 1.941   
## Initial\_adminElective Admission 1.512 0.340 0.025 4.449 0.000 0.846 2.178   
## HighBloodYes 0.972 0.786 0.018 1.237 0.216 -0.568 2.512   
## StrokeYes -0.117 0.347 -0.002 -0.337 0.736 -0.797 0.563   
## Complication\_riskLow 1.216 0.363 0.019 3.347 0.001 0.504 1.928   
## Complication\_riskHigh 0.359 0.316 0.006 1.134 0.257 -0.261 0.979   
## OverweightYes -0.194 0.304 -0.003 -0.638 0.524 -0.790 0.402   
## ArthritisYes 0.662 0.288 0.012 2.296 0.022 0.097 1.227   
## HyperlipidemiaYes -0.396 0.292 -0.007 -1.357 0.175 -0.969 0.176   
## BackPainYes 0.324 0.281 0.006 1.155 0.248 -0.226 0.875   
## AnxietyYes 0.553 0.296 0.010 1.869 0.062 -0.027 1.132   
## Allergic\_rhinitisYes 0.398 0.283 0.007 1.409 0.159 -0.156 0.952   
## Reflux\_esophagitisYes 0.402 0.280 0.008 1.432 0.152 -0.148 0.951   
## AsthmaYes 0.074 0.305 0.001 0.244 0.807 -0.523 0.671   
## ServicesCT Scan -1.057 0.438 -0.013 -2.414 0.016 -1.915 -0.198   
## ServicesIntravenous -0.089 0.312 -0.002 -0.287 0.774 -0.700 0.521   
## ServicesMRI -0.162 0.733 -0.001 -0.221 0.825 -1.599 1.275   
## Additional\_charges 0.000 0.000 -0.040 -1.883 0.060 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Asthma   
##   
## Backward Elimination: Step 3   
##   
## Variable Asthma Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.787   
## R-Squared 0.726 Coef. Var 40.013   
## Adj. R-Squared 0.725 MSE 190.074   
## Pred R-Squared 0.724 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026650.163 32 157082.818 826.428 0.0000   
## Residual 1894472.058 9967 190.074   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.432 1.584 11.005 0.000 14.327 20.537   
## AreaSuburban 0.152 0.337 0.003 0.449 0.653 -0.510 0.813   
## AreaUrban 0.341 0.338 0.006 1.008 0.314 -0.322 1.004   
## Children 0.039 0.064 0.003 0.616 0.538 -0.086 0.164   
## Age 0.040 0.020 0.031 1.949 0.051 0.000 0.079   
## Income 0.000 0.000 -0.004 -0.693 0.488 0.000 0.000   
## MaritalMarried -0.035 0.438 -0.001 -0.079 0.937 -0.892 0.823   
## MaritalNever Married 0.438 0.440 0.007 0.995 0.320 -0.424 1.300   
## MaritalSeparated 0.786 0.439 0.012 1.789 0.074 -0.075 1.647   
## MaritalWidowed 0.285 0.436 0.004 0.654 0.513 -0.570 1.141   
## ReAdmisYes 46.461 0.287 0.851 162.136 0.000 45.899 47.023   
## VitD\_levels -0.079 0.068 -0.006 -1.151 0.250 -0.213 0.055   
## Doc\_visits -0.167 0.132 -0.007 -1.264 0.206 -0.426 0.092   
## Full\_meals\_eaten -0.163 0.137 -0.006 -1.191 0.234 -0.432 0.105   
## vitD\_supp 0.267 0.220 0.006 1.214 0.225 -0.164 0.697   
## Soft\_drinkYes -0.180 0.316 -0.003 -0.571 0.568 -0.800 0.439   
## Initial\_adminObservation Admission 1.267 0.344 0.021 3.682 0.000 0.592 1.941   
## Initial\_adminElective Admission 1.512 0.340 0.025 4.451 0.000 0.846 2.178   
## HighBloodYes 0.970 0.786 0.018 1.234 0.217 -0.570 2.510   
## StrokeYes -0.117 0.347 -0.002 -0.337 0.736 -0.797 0.563   
## Complication\_riskLow 1.216 0.363 0.019 3.347 0.001 0.504 1.928   
## Complication\_riskHigh 0.358 0.316 0.006 1.131 0.258 -0.262 0.978   
## OverweightYes -0.193 0.304 -0.003 -0.635 0.526 -0.789 0.403   
## ArthritisYes 0.661 0.288 0.012 2.294 0.022 0.096 1.226   
## HyperlipidemiaYes -0.397 0.292 -0.007 -1.359 0.174 -0.970 0.176   
## BackPainYes 0.325 0.281 0.006 1.158 0.247 -0.225 0.876   
## AnxietyYes 0.553 0.296 0.010 1.872 0.061 -0.026 1.133   
## Allergic\_rhinitisYes 0.399 0.283 0.007 1.410 0.158 -0.155 0.952   
## Reflux\_esophagitisYes 0.401 0.280 0.008 1.432 0.152 -0.148 0.951   
## ServicesCT Scan -1.056 0.438 -0.013 -2.411 0.016 -1.914 -0.197   
## ServicesIntravenous -0.090 0.312 -0.002 -0.290 0.772 -0.701 0.521   
## ServicesMRI -0.163 0.733 -0.001 -0.222 0.825 -1.600 1.274   
## Additional\_charges 0.000 0.000 -0.040 -1.879 0.060 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Stroke   
##   
## Backward Elimination: Step 4   
##   
## Variable Stroke Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.786   
## R-Squared 0.726 Coef. Var 40.012   
## Adj. R-Squared 0.725 MSE 190.058   
## Pred R-Squared 0.725 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026628.623 31 162149.310 853.159 0.0000   
## Residual 1894493.598 9968 190.058   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.405 1.582 11.002 0.000 14.304 20.505   
## AreaSuburban 0.152 0.337 0.003 0.451 0.652 -0.509 0.813   
## AreaUrban 0.341 0.338 0.006 1.007 0.314 -0.322 1.004   
## Children 0.039 0.064 0.003 0.615 0.539 -0.086 0.164   
## Age 0.040 0.020 0.031 1.983 0.047 0.000 0.080   
## Income 0.000 0.000 -0.004 -0.694 0.488 0.000 0.000   
## MaritalMarried -0.033 0.438 -0.001 -0.075 0.940 -0.891 0.825   
## MaritalNever Married 0.439 0.440 0.007 0.999 0.318 -0.423 1.301   
## MaritalSeparated 0.787 0.439 0.012 1.791 0.073 -0.075 1.648   
## MaritalWidowed 0.287 0.436 0.004 0.659 0.510 -0.568 1.143   
## ReAdmisYes 46.461 0.287 0.851 162.143 0.000 45.899 47.023   
## VitD\_levels -0.079 0.068 -0.006 -1.154 0.248 -0.213 0.055   
## Doc\_visits -0.167 0.132 -0.007 -1.263 0.207 -0.426 0.092   
## Full\_meals\_eaten -0.163 0.137 -0.006 -1.192 0.233 -0.432 0.105   
## vitD\_supp 0.266 0.220 0.006 1.212 0.225 -0.164 0.697   
## Soft\_drinkYes -0.181 0.316 -0.003 -0.572 0.567 -0.800 0.438   
## Initial\_adminObservation Admission 1.264 0.344 0.021 3.676 0.000 0.590 1.939   
## Initial\_adminElective Admission 1.510 0.340 0.025 4.446 0.000 0.844 2.176   
## HighBloodYes 0.991 0.783 0.019 1.266 0.206 -0.544 2.526   
## Complication\_riskLow 1.216 0.363 0.019 3.347 0.001 0.504 1.928   
## Complication\_riskHigh 0.359 0.316 0.006 1.135 0.256 -0.261 0.979   
## OverweightYes -0.193 0.304 -0.003 -0.634 0.526 -0.789 0.403   
## ArthritisYes 0.663 0.288 0.012 2.300 0.021 0.098 1.228   
## HyperlipidemiaYes -0.396 0.292 -0.007 -1.355 0.176 -0.968 0.177   
## BackPainYes 0.325 0.281 0.006 1.157 0.247 -0.226 0.876   
## AnxietyYes 0.555 0.296 0.010 1.878 0.060 -0.024 1.134   
## Allergic\_rhinitisYes 0.399 0.283 0.007 1.412 0.158 -0.155 0.953   
## Reflux\_esophagitisYes 0.401 0.280 0.008 1.432 0.152 -0.148 0.951   
## ServicesCT Scan -1.057 0.438 -0.013 -2.414 0.016 -1.915 -0.199   
## ServicesIntravenous -0.088 0.312 -0.002 -0.284 0.777 -0.699 0.522   
## ServicesMRI -0.161 0.733 -0.001 -0.220 0.826 -1.598 1.276   
## Additional\_charges 0.000 0.000 -0.040 -1.917 0.055 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Area   
##   
## Backward Elimination: Step 5   
##   
## Variable Area Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.010   
## Adj. R-Squared 0.725 MSE 190.039   
## Pred R-Squared 0.725 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026435.220 29 173325.352 912.052 0.0000   
## Residual 1894687.001 9970 190.039   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.520 1.575 11.126 0.000 14.433 20.607   
## Children 0.039 0.064 0.003 0.609 0.542 -0.086 0.164   
## Age 0.040 0.020 0.032 1.997 0.046 0.001 0.080   
## Income 0.000 0.000 -0.004 -0.684 0.494 0.000 0.000   
## MaritalMarried -0.032 0.438 0.000 -0.072 0.942 -0.889 0.826   
## MaritalNever Married 0.438 0.440 0.007 0.996 0.319 -0.424 1.300   
## MaritalSeparated 0.785 0.439 0.012 1.787 0.074 -0.076 1.646   
## MaritalWidowed 0.289 0.436 0.004 0.661 0.508 -0.567 1.144   
## ReAdmisYes 46.464 0.287 0.851 162.166 0.000 45.902 47.025   
## VitD\_levels -0.077 0.068 -0.006 -1.121 0.262 -0.211 0.057   
## Doc\_visits -0.167 0.132 -0.007 -1.265 0.206 -0.426 0.092   
## Full\_meals\_eaten -0.164 0.137 -0.006 -1.195 0.232 -0.432 0.105   
## vitD\_supp 0.268 0.220 0.006 1.218 0.223 -0.163 0.698   
## Soft\_drinkYes -0.183 0.316 -0.003 -0.580 0.562 -0.802 0.436   
## Initial\_adminObservation Admission 1.264 0.344 0.021 3.677 0.000 0.590 1.939   
## Initial\_adminElective Admission 1.504 0.340 0.025 4.429 0.000 0.839 2.170   
## HighBloodYes 1.008 0.783 0.019 1.288 0.198 -0.526 2.543   
## Complication\_riskLow 1.218 0.363 0.019 3.352 0.001 0.506 1.930   
## Complication\_riskHigh 0.360 0.316 0.006 1.139 0.255 -0.260 0.980   
## OverweightYes -0.194 0.304 -0.003 -0.639 0.523 -0.791 0.402   
## ArthritisYes 0.661 0.288 0.012 2.295 0.022 0.097 1.226   
## HyperlipidemiaYes -0.393 0.292 -0.007 -1.346 0.178 -0.966 0.179   
## BackPainYes 0.323 0.281 0.006 1.150 0.250 -0.228 0.873   
## AnxietyYes 0.563 0.295 0.010 1.907 0.057 -0.016 1.143   
## Allergic\_rhinitisYes 0.404 0.283 0.008 1.429 0.153 -0.150 0.958   
## Reflux\_esophagitisYes 0.404 0.280 0.008 1.439 0.150 -0.146 0.953   
## ServicesCT Scan -1.059 0.438 -0.013 -2.420 0.016 -1.917 -0.201   
## ServicesIntravenous -0.087 0.311 -0.002 -0.281 0.779 -0.698 0.523   
## ServicesMRI -0.167 0.733 -0.001 -0.228 0.820 -1.604 1.270   
## Additional\_charges 0.000 0.000 -0.041 -1.935 0.053 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Soft\_drink   
##   
## Backward Elimination: Step 6   
##   
## Variable Soft\_drink Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.008   
## Adj. R-Squared 0.725 MSE 190.026   
## Pred R-Squared 0.725 MAE 10.207   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026371.271 28 179513.260 944.677 0.0000   
## Residual 1894750.950 9971 190.026   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.480 1.573 11.112 0.000 14.396 20.564   
## Children 0.039 0.064 0.003 0.605 0.545 -0.086 0.164   
## Age 0.041 0.020 0.032 2.001 0.045 0.001 0.080   
## Income 0.000 0.000 -0.004 -0.686 0.493 0.000 0.000   
## MaritalMarried -0.033 0.438 -0.001 -0.075 0.940 -0.891 0.825   
## MaritalNever Married 0.435 0.440 0.007 0.990 0.322 -0.426 1.297   
## MaritalSeparated 0.785 0.439 0.012 1.788 0.074 -0.075 1.646   
## MaritalWidowed 0.291 0.436 0.004 0.666 0.505 -0.565 1.146   
## ReAdmisYes 46.462 0.287 0.851 162.171 0.000 45.901 47.024   
## VitD\_levels -0.077 0.068 -0.006 -1.123 0.262 -0.211 0.057   
## Doc\_visits -0.168 0.132 -0.007 -1.272 0.203 -0.427 0.091   
## Full\_meals\_eaten -0.165 0.137 -0.006 -1.208 0.227 -0.434 0.103   
## vitD\_supp 0.270 0.220 0.006 1.227 0.220 -0.161 0.700   
## Initial\_adminObservation Admission 1.269 0.344 0.021 3.692 0.000 0.595 1.943   
## Initial\_adminElective Admission 1.506 0.340 0.025 4.435 0.000 0.840 2.172   
## HighBloodYes 1.012 0.783 0.019 1.293 0.196 -0.522 2.546   
## Complication\_riskLow 1.219 0.363 0.019 3.356 0.001 0.507 1.931   
## Complication\_riskHigh 0.361 0.316 0.006 1.143 0.253 -0.259 0.981   
## OverweightYes -0.194 0.304 -0.003 -0.637 0.524 -0.790 0.402   
## ArthritisYes 0.663 0.288 0.012 2.300 0.021 0.098 1.228   
## HyperlipidemiaYes -0.396 0.292 -0.007 -1.356 0.175 -0.968 0.176   
## BackPainYes 0.320 0.281 0.006 1.140 0.254 -0.230 0.870   
## AnxietyYes 0.561 0.295 0.010 1.898 0.058 -0.018 1.140   
## Allergic\_rhinitisYes 0.406 0.282 0.008 1.439 0.150 -0.147 0.960   
## Reflux\_esophagitisYes 0.404 0.280 0.008 1.441 0.150 -0.145 0.953   
## ServicesCT Scan -1.058 0.438 -0.013 -2.417 0.016 -1.916 -0.200   
## ServicesIntravenous -0.088 0.311 -0.002 -0.283 0.777 -0.699 0.522   
## ServicesMRI -0.173 0.733 -0.001 -0.235 0.814 -1.609 1.264   
## Additional\_charges 0.000 0.000 -0.041 -1.939 0.052 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Children   
##   
## Backward Elimination: Step 7   
##   
## Variable Children Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.007   
## Adj. R-Squared 0.725 MSE 190.014   
## Pred R-Squared 0.725 MAE 10.207   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------------  
## Regression 5026301.773 27 186159.325 979.713 0.0000   
## Residual 1894820.448 9972 190.014   
## Total 6921122.221 9999   
## --------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.558 1.568 11.198 0.000 14.484 20.631   
## Age 0.040 0.020 0.032 1.992 0.046 0.001 0.080   
## Income 0.000 0.000 -0.004 -0.682 0.495 0.000 0.000   
## MaritalMarried -0.032 0.438 0.000 -0.074 0.941 -0.890 0.825   
## MaritalNever Married 0.430 0.440 0.007 0.978 0.328 -0.432 1.292   
## MaritalSeparated 0.781 0.439 0.012 1.778 0.075 -0.080 1.642   
## MaritalWidowed 0.287 0.436 0.004 0.658 0.511 -0.568 1.142   
## ReAdmisYes 46.466 0.286 0.851 162.236 0.000 45.905 47.028   
## VitD\_levels -0.076 0.068 -0.006 -1.117 0.264 -0.211 0.058   
## Doc\_visits -0.168 0.132 -0.007 -1.274 0.203 -0.427 0.091   
## Full\_meals\_eaten -0.165 0.137 -0.006 -1.206 0.228 -0.434 0.103   
## vitD\_supp 0.269 0.220 0.006 1.225 0.221 -0.162 0.699   
## Initial\_adminObservation Admission 1.268 0.344 0.021 3.689 0.000 0.594 1.942   
## Initial\_adminElective Admission 1.505 0.340 0.025 4.433 0.000 0.840 2.171   
## HighBloodYes 1.004 0.783 0.019 1.283 0.200 -0.530 2.538   
## Complication\_riskLow 1.220 0.363 0.019 3.359 0.001 0.508 1.932   
## Complication\_riskHigh 0.362 0.316 0.007 1.145 0.252 -0.258 0.982   
## OverweightYes -0.196 0.304 -0.003 -0.645 0.519 -0.792 0.400   
## ArthritisYes 0.664 0.288 0.012 2.305 0.021 0.099 1.229   
## HyperlipidemiaYes -0.396 0.292 -0.007 -1.358 0.175 -0.969 0.176   
## BackPainYes 0.319 0.281 0.006 1.135 0.257 -0.232 0.869   
## AnxietyYes 0.562 0.295 0.010 1.903 0.057 -0.017 1.141   
## Allergic\_rhinitisYes 0.403 0.282 0.007 1.427 0.154 -0.151 0.957   
## Reflux\_esophagitisYes 0.405 0.280 0.008 1.445 0.149 -0.144 0.954   
## ServicesCT Scan -1.057 0.438 -0.013 -2.415 0.016 -1.915 -0.199   
## ServicesIntravenous -0.086 0.311 -0.002 -0.277 0.782 -0.697 0.524   
## ServicesMRI -0.174 0.733 -0.001 -0.238 0.812 -1.611 1.262   
## Additional\_charges 0.000 0.000 -0.041 -1.928 0.054 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Overweight   
##   
## Backward Elimination: Step 8   
##   
## Variable Overweight Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.784   
## R-Squared 0.726 Coef. Var 40.006   
## Adj. R-Squared 0.726 MSE 190.003   
## Pred R-Squared 0.725 MAE 10.207   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5026222.808 26 193316.262 1017.438 0.0000   
## Residual 1894899.413 9973 190.003   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.427 1.555 11.209 0.000 14.380 20.475   
## Age 0.040 0.020 0.032 2.000 0.045 0.001 0.080   
## Income 0.000 0.000 -0.004 -0.670 0.503 0.000 0.000   
## MaritalMarried -0.035 0.438 -0.001 -0.081 0.935 -0.893 0.822   
## MaritalNever Married 0.425 0.439 0.006 0.967 0.333 -0.436 1.286   
## MaritalSeparated 0.779 0.439 0.012 1.775 0.076 -0.082 1.640   
## MaritalWidowed 0.283 0.436 0.004 0.648 0.517 -0.572 1.138   
## ReAdmisYes 46.468 0.286 0.851 162.254 0.000 45.907 47.029   
## VitD\_levels -0.077 0.068 -0.006 -1.124 0.261 -0.211 0.057   
## Doc\_visits -0.169 0.132 -0.007 -1.282 0.200 -0.428 0.090   
## Full\_meals\_eaten -0.164 0.137 -0.006 -1.200 0.230 -0.433 0.104   
## vitD\_supp 0.270 0.220 0.006 1.228 0.220 -0.161 0.700   
## Initial\_adminObservation Admission 1.266 0.344 0.021 3.682 0.000 0.592 1.940   
## Initial\_adminElective Admission 1.504 0.340 0.025 4.428 0.000 0.838 2.169   
## HighBloodYes 1.004 0.783 0.019 1.282 0.200 -0.530 2.538   
## Complication\_riskLow 1.223 0.363 0.019 3.368 0.001 0.511 1.935   
## Complication\_riskHigh 0.366 0.316 0.007 1.156 0.248 -0.254 0.985   
## ArthritisYes 0.663 0.288 0.012 2.303 0.021 0.099 1.228   
## HyperlipidemiaYes -0.395 0.292 -0.007 -1.354 0.176 -0.968 0.177   
## BackPainYes 0.317 0.281 0.006 1.127 0.260 -0.234 0.867   
## AnxietyYes 0.564 0.295 0.010 1.911 0.056 -0.015 1.143   
## Allergic\_rhinitisYes 0.402 0.282 0.007 1.425 0.154 -0.151 0.956   
## Reflux\_esophagitisYes 0.407 0.280 0.008 1.453 0.146 -0.142 0.957   
## ServicesCT Scan -1.058 0.438 -0.013 -2.417 0.016 -1.916 -0.200   
## ServicesIntravenous -0.087 0.311 -0.002 -0.279 0.780 -0.698 0.523   
## ServicesMRI -0.173 0.733 -0.001 -0.236 0.813 -1.610 1.263   
## Additional\_charges 0.000 0.000 -0.041 -1.935 0.053 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Income   
##   
## Backward Elimination: Step 9   
##   
## Variable Income Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.784   
## R-Squared 0.726 Coef. Var 40.005   
## Adj. R-Squared 0.726 MSE 189.992   
## Pred R-Squared 0.725 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5026137.608 25 201045.504 1058.176 0.0000   
## Residual 1894984.613 9974 189.992   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.290 1.541 11.219 0.000 14.269 20.311   
## Age 0.041 0.020 0.032 2.005 0.045 0.001 0.080   
## MaritalMarried -0.031 0.437 0.000 -0.071 0.943 -0.888 0.826   
## MaritalNever Married 0.425 0.439 0.006 0.968 0.333 -0.436 1.287   
## MaritalSeparated 0.783 0.439 0.012 1.782 0.075 -0.078 1.643   
## MaritalWidowed 0.287 0.436 0.004 0.657 0.511 -0.568 1.142   
## ReAdmisYes 46.470 0.286 0.851 162.274 0.000 45.909 47.031   
## VitD\_levels -0.076 0.068 -0.006 -1.116 0.264 -0.210 0.058   
## Doc\_visits -0.170 0.132 -0.007 -1.291 0.197 -0.429 0.088   
## Full\_meals\_eaten -0.163 0.137 -0.006 -1.193 0.233 -0.432 0.105   
## vitD\_supp 0.269 0.220 0.006 1.227 0.220 -0.161 0.700   
## Initial\_adminObservation Admission 1.260 0.344 0.021 3.667 0.000 0.587 1.934   
## Initial\_adminElective Admission 1.501 0.340 0.025 4.421 0.000 0.836 2.167   
## HighBloodYes 1.005 0.783 0.019 1.284 0.199 -0.529 2.539   
## Complication\_riskLow 1.221 0.363 0.019 3.363 0.001 0.510 1.933   
## Complication\_riskHigh 0.365 0.316 0.007 1.153 0.249 -0.255 0.984   
## ArthritisYes 0.664 0.288 0.012 2.306 0.021 0.100 1.229   
## HyperlipidemiaYes -0.397 0.292 -0.007 -1.360 0.174 -0.969 0.175   
## BackPainYes 0.315 0.281 0.006 1.121 0.262 -0.236 0.865   
## AnxietyYes 0.564 0.295 0.010 1.911 0.056 -0.015 1.143   
## Allergic\_rhinitisYes 0.403 0.282 0.007 1.425 0.154 -0.151 0.956   
## Reflux\_esophagitisYes 0.404 0.280 0.008 1.442 0.149 -0.145 0.953   
## ServicesCT Scan -1.056 0.438 -0.013 -2.413 0.016 -1.914 -0.198   
## ServicesIntravenous -0.088 0.311 -0.002 -0.283 0.777 -0.699 0.522   
## ServicesMRI -0.172 0.733 -0.001 -0.235 0.814 -1.609 1.264   
## Additional\_charges 0.000 0.000 -0.041 -1.936 0.053 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Marital   
##   
## Backward Elimination: Step 10   
##   
## Variable Marital Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.784   
## R-Squared 0.726 Coef. Var 40.006   
## Adj. R-Squared 0.725 MSE 190.005   
## Pred R-Squared 0.725 MAE 10.208   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5025247.408 21 239297.496 1259.424 0.0000   
## Residual 1895874.813 9978 190.005   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.616 1.513 11.643 0.000 14.650 20.582   
## Age 0.041 0.020 0.032 2.002 0.045 0.001 0.080   
## ReAdmisYes 46.474 0.286 0.851 162.325 0.000 45.913 47.035   
## VitD\_levels -0.077 0.068 -0.006 -1.124 0.261 -0.211 0.057   
## Doc\_visits -0.172 0.132 -0.007 -1.306 0.191 -0.431 0.086   
## Full\_meals\_eaten -0.166 0.137 -0.006 -1.215 0.224 -0.435 0.102   
## vitD\_supp 0.271 0.220 0.006 1.236 0.217 -0.159 0.702   
## Initial\_adminObservation Admission 1.266 0.344 0.021 3.685 0.000 0.593 1.940   
## Initial\_adminElective Admission 1.507 0.339 0.025 4.439 0.000 0.841 2.172   
## HighBloodYes 0.997 0.783 0.019 1.274 0.203 -0.537 2.531   
## Complication\_riskLow 1.227 0.363 0.019 3.380 0.001 0.516 1.939   
## Complication\_riskHigh 0.362 0.316 0.006 1.144 0.252 -0.258 0.981   
## ArthritisYes 0.669 0.288 0.012 2.325 0.020 0.105 1.234   
## HyperlipidemiaYes -0.394 0.292 -0.007 -1.349 0.177 -0.966 0.178   
## BackPainYes 0.303 0.281 0.006 1.082 0.279 -0.246 0.853   
## AnxietyYes 0.565 0.295 0.010 1.914 0.056 -0.014 1.144   
## Allergic\_rhinitisYes 0.396 0.282 0.007 1.404 0.160 -0.157 0.950   
## Reflux\_esophagitisYes 0.409 0.280 0.008 1.459 0.145 -0.140 0.958   
## ServicesCT Scan -1.061 0.438 -0.013 -2.425 0.015 -1.919 -0.203   
## ServicesIntravenous -0.094 0.311 -0.002 -0.303 0.762 -0.705 0.516   
## ServicesMRI -0.185 0.733 -0.001 -0.252 0.801 -1.621 1.251   
## Additional\_charges 0.000 0.000 -0.041 -1.941 0.052 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - BackPain   
##   
## Backward Elimination: Step 11   
##   
## Variable BackPain Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.784   
## R-Squared 0.726 Coef. Var 40.007   
## Adj. R-Squared 0.725 MSE 190.009   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5025025.060 20 251251.253 1322.314 0.0000   
## Residual 1896097.161 9979 190.009   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 17.735 1.509 11.752 0.000 14.777 20.693   
## Age 0.041 0.020 0.032 2.020 0.043 0.001 0.081   
## ReAdmisYes 46.478 0.286 0.851 162.349 0.000 45.917 47.039   
## VitD\_levels -0.078 0.068 -0.006 -1.134 0.257 -0.212 0.057   
## Doc\_visits -0.171 0.132 -0.007 -1.298 0.194 -0.430 0.087   
## Full\_meals\_eaten -0.169 0.137 -0.006 -1.232 0.218 -0.437 0.100   
## vitD\_supp 0.271 0.220 0.006 1.233 0.218 -0.160 0.701   
## Initial\_adminObservation Admission 1.268 0.344 0.021 3.691 0.000 0.595 1.942   
## Initial\_adminElective Admission 1.504 0.339 0.025 4.429 0.000 0.838 2.169   
## HighBloodYes 1.006 0.782 0.019 1.285 0.199 -0.528 2.540   
## Complication\_riskLow 1.235 0.363 0.019 3.400 0.001 0.523 1.946   
## Complication\_riskHigh 0.362 0.316 0.007 1.146 0.252 -0.257 0.982   
## ArthritisYes 0.663 0.288 0.012 2.304 0.021 0.099 1.228   
## HyperlipidemiaYes -0.394 0.292 -0.007 -1.350 0.177 -0.966 0.178   
## AnxietyYes 0.568 0.295 0.010 1.924 0.054 -0.011 1.147   
## Allergic\_rhinitisYes 0.398 0.282 0.007 1.408 0.159 -0.156 0.951   
## Reflux\_esophagitisYes 0.414 0.280 0.008 1.477 0.140 -0.135 0.963   
## ServicesCT Scan -1.055 0.438 -0.013 -2.411 0.016 -1.913 -0.197   
## ServicesIntravenous -0.097 0.311 -0.002 -0.313 0.754 -0.708 0.513   
## ServicesMRI -0.186 0.733 -0.001 -0.254 0.800 -1.622 1.250   
## Additional\_charges 0.000 0.000 -0.041 -1.952 0.051 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - VitD\_levels   
##   
## Backward Elimination: Step 12   
##   
## Variable VitD\_levels Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.007   
## Adj. R-Squared 0.725 MSE 190.014   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5024780.759 19 264462.145 1391.802 0.0000   
## Residual 1896341.462 9980 190.014   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.350 0.886 18.444 0.000 14.613 18.088   
## Age 0.040 0.020 0.032 2.000 0.045 0.001 0.080   
## ReAdmisYes 46.477 0.286 0.851 162.344 0.000 45.916 47.038   
## Doc\_visits -0.173 0.132 -0.007 -1.308 0.191 -0.431 0.086   
## Full\_meals\_eaten -0.172 0.137 -0.007 -1.258 0.209 -0.441 0.096   
## vitD\_supp 0.272 0.220 0.007 1.241 0.215 -0.158 0.703   
## Initial\_adminObservation Admission 1.277 0.344 0.021 3.719 0.000 0.604 1.951   
## Initial\_adminElective Admission 1.513 0.339 0.025 4.459 0.000 0.848 2.178   
## HighBloodYes 0.992 0.782 0.019 1.267 0.205 -0.542 2.525   
## Complication\_riskLow 1.239 0.363 0.019 3.414 0.001 0.528 1.951   
## Complication\_riskHigh 0.364 0.316 0.007 1.152 0.249 -0.255 0.984   
## ArthritisYes 0.664 0.288 0.012 2.305 0.021 0.099 1.228   
## HyperlipidemiaYes -0.390 0.292 -0.007 -1.336 0.182 -0.962 0.182   
## AnxietyYes 0.566 0.295 0.010 1.916 0.055 -0.013 1.145   
## Allergic\_rhinitisYes 0.400 0.282 0.007 1.415 0.157 -0.154 0.953   
## Reflux\_esophagitisYes 0.418 0.280 0.008 1.493 0.135 -0.131 0.967   
## ServicesCT Scan -1.058 0.438 -0.013 -2.417 0.016 -1.916 -0.200   
## ServicesIntravenous -0.097 0.311 -0.002 -0.310 0.756 -0.707 0.514   
## ServicesMRI -0.176 0.733 -0.001 -0.240 0.810 -1.612 1.260   
## Additional\_charges 0.000 0.000 -0.041 -1.935 0.053 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - vitD\_supp   
##   
## Backward Elimination: Step 13   
##   
## Variable vitD\_supp Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.008   
## Adj. R-Squared 0.725 MSE 190.024   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5024488.331 18 279138.241 1468.96 0.0000   
## Residual 1896633.890 9981 190.024   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.444 0.883 18.617 0.000 14.713 18.175   
## Age 0.041 0.020 0.032 2.008 0.045 0.001 0.080   
## ReAdmisYes 46.481 0.286 0.852 162.363 0.000 45.920 47.042   
## Doc\_visits -0.172 0.132 -0.007 -1.300 0.193 -0.430 0.087   
## Full\_meals\_eaten -0.176 0.137 -0.007 -1.283 0.200 -0.444 0.093   
## Initial\_adminObservation Admission 1.272 0.344 0.021 3.702 0.000 0.598 1.945   
## Initial\_adminElective Admission 1.520 0.339 0.025 4.479 0.000 0.855 2.185   
## HighBloodYes 0.997 0.782 0.019 1.274 0.203 -0.537 2.530   
## Complication\_riskLow 1.242 0.363 0.019 3.421 0.001 0.530 1.954   
## Complication\_riskHigh 0.371 0.316 0.007 1.172 0.241 -0.249 0.990   
## ArthritisYes 0.659 0.288 0.012 2.290 0.022 0.095 1.223   
## HyperlipidemiaYes -0.385 0.292 -0.007 -1.320 0.187 -0.957 0.187   
## AnxietyYes 0.566 0.295 0.010 1.916 0.055 -0.013 1.145   
## Allergic\_rhinitisYes 0.403 0.282 0.007 1.426 0.154 -0.151 0.956   
## Reflux\_esophagitisYes 0.416 0.280 0.008 1.487 0.137 -0.133 0.966   
## ServicesCT Scan -1.056 0.438 -0.013 -2.414 0.016 -1.914 -0.198   
## ServicesIntravenous -0.091 0.311 -0.002 -0.293 0.770 -0.701 0.519   
## ServicesMRI -0.165 0.733 -0.001 -0.225 0.822 -1.601 1.271   
## Additional\_charges 0.000 0.000 -0.041 -1.939 0.053 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - HighBlood   
##   
## Backward Elimination: Step 14   
##   
## Variable HighBlood Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.785   
## R-Squared 0.726 Coef. Var 40.009   
## Adj. R-Squared 0.725 MSE 190.036   
## Pred R-Squared 0.725 MAE 10.210   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5024180.018 17 295540.001 1555.177 0.0000   
## Residual 1896942.203 9982 190.036   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.739 0.852 19.635 0.000 15.068 18.410   
## Age 0.018 0.010 0.014 1.869 0.062 -0.001 0.037   
## ReAdmisYes 46.481 0.286 0.852 162.357 0.000 45.919 47.042   
## Doc\_visits -0.169 0.132 -0.007 -1.282 0.200 -0.428 0.089   
## Full\_meals\_eaten -0.176 0.137 -0.007 -1.288 0.198 -0.445 0.092   
## Initial\_adminObservation Admission 1.329 0.341 0.022 3.901 0.000 0.661 1.996   
## Initial\_adminElective Admission 1.566 0.337 0.026 4.641 0.000 0.904 2.227   
## Complication\_riskLow 1.251 0.363 0.019 3.445 0.001 0.539 1.962   
## Complication\_riskHigh 0.332 0.315 0.006 1.055 0.292 -0.285 0.949   
## ArthritisYes 0.666 0.288 0.012 2.314 0.021 0.102 1.230   
## HyperlipidemiaYes -0.387 0.292 -0.007 -1.326 0.185 -0.959 0.185   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## Allergic\_rhinitisYes 0.405 0.282 0.008 1.433 0.152 -0.149 0.958   
## Reflux\_esophagitisYes 0.414 0.280 0.008 1.479 0.139 -0.135 0.963   
## ServicesCT Scan -1.052 0.438 -0.013 -2.403 0.016 -1.909 -0.194   
## ServicesIntravenous -0.094 0.311 -0.002 -0.303 0.762 -0.704 0.516   
## ServicesMRI -0.175 0.733 -0.001 -0.239 0.811 -1.611 1.261   
## Additional\_charges 0.000 0.000 -0.016 -2.091 0.037 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Doc\_visits   
##   
## Backward Elimination: Step 15   
##   
## Variable Doc\_visits Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.786   
## R-Squared 0.726 Coef. Var 40.011   
## Adj. R-Squared 0.725 MSE 190.049   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5023867.502 16 313991.719 1652.166 0.0000   
## Residual 1897254.719 9983 190.049   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.889 0.537 29.599 0.000 14.837 16.942   
## Age 0.018 0.010 0.014 1.865 0.062 -0.001 0.037   
## ReAdmisYes 46.481 0.286 0.852 162.352 0.000 45.919 47.042   
## Full\_meals\_eaten -0.176 0.137 -0.007 -1.283 0.199 -0.444 0.093   
## Initial\_adminObservation Admission 1.324 0.341 0.022 3.888 0.000 0.657 1.992   
## Initial\_adminElective Admission 1.573 0.337 0.026 4.662 0.000 0.912 2.234   
## Complication\_riskLow 1.251 0.363 0.019 3.446 0.001 0.539 1.963   
## Complication\_riskHigh 0.327 0.315 0.006 1.039 0.299 -0.290 0.944   
## ArthritisYes 0.666 0.288 0.012 2.315 0.021 0.102 1.231   
## HyperlipidemiaYes -0.377 0.292 -0.007 -1.292 0.196 -0.949 0.195   
## AnxietyYes 0.565 0.295 0.010 1.914 0.056 -0.014 1.144   
## Allergic\_rhinitisYes 0.404 0.282 0.007 1.429 0.153 -0.150 0.957   
## Reflux\_esophagitisYes 0.416 0.280 0.008 1.486 0.137 -0.133 0.965   
## ServicesCT Scan -1.058 0.438 -0.013 -2.418 0.016 -1.916 -0.200   
## ServicesIntravenous -0.091 0.311 -0.002 -0.293 0.770 -0.701 0.519   
## ServicesMRI -0.163 0.733 -0.001 -0.223 0.824 -1.599 1.273   
## Additional\_charges 0.000 0.000 -0.016 -2.095 0.036 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Full\_meals\_eaten   
##   
## Backward Elimination: Step 16   
##   
## Variable Full\_meals\_eaten Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.786   
## R-Squared 0.726 Coef. Var 40.012   
## Adj. R-Squared 0.725 MSE 190.061   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5023554.495 15 334903.633 1762.086 0.0000   
## Residual 1897567.726 9984 190.061   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.722 0.521 30.194 0.000 14.701 16.742   
## Age 0.018 0.010 0.014 1.873 0.061 -0.001 0.037   
## ReAdmisYes 46.485 0.286 0.852 162.374 0.000 45.924 47.046   
## Initial\_adminObservation Admission 1.324 0.341 0.022 3.886 0.000 0.656 1.991   
## Initial\_adminElective Admission 1.578 0.337 0.026 4.676 0.000 0.916 2.239   
## Complication\_riskLow 1.254 0.363 0.020 3.455 0.001 0.543 1.966   
## Complication\_riskHigh 0.323 0.315 0.006 1.027 0.305 -0.294 0.940   
## ArthritisYes 0.662 0.288 0.012 2.300 0.021 0.098 1.226   
## HyperlipidemiaYes -0.377 0.292 -0.007 -1.293 0.196 -0.949 0.195   
## AnxietyYes 0.562 0.295 0.010 1.904 0.057 -0.017 1.141   
## Allergic\_rhinitisYes 0.398 0.282 0.007 1.410 0.159 -0.155 0.952   
## Reflux\_esophagitisYes 0.416 0.280 0.008 1.486 0.137 -0.133 0.965   
## ServicesCT Scan -1.057 0.438 -0.013 -2.416 0.016 -1.915 -0.200   
## ServicesIntravenous -0.096 0.311 -0.002 -0.309 0.757 -0.706 0.514   
## ServicesMRI -0.147 0.733 -0.001 -0.201 0.841 -1.583 1.289   
## Additional\_charges 0.000 0.000 -0.016 -2.117 0.034 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Hyperlipidemia   
##   
## Backward Elimination: Step 17   
##   
## Variable Hyperlipidemia Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.787   
## R-Squared 0.726 Coef. Var 40.013   
## Adj. R-Squared 0.725 MSE 190.074   
## Pred R-Squared 0.725 MAE 10.209   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5023236.747 14 358802.625 1887.703 0.0000   
## Residual 1897885.475 9985 190.074   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.587 0.510 30.554 0.000 14.587 16.587   
## Age 0.018 0.010 0.014 1.862 0.063 -0.001 0.037   
## ReAdmisYes 46.484 0.286 0.852 162.365 0.000 45.922 47.045   
## Initial\_adminObservation Admission 1.330 0.341 0.022 3.905 0.000 0.663 1.998   
## Initial\_adminElective Admission 1.585 0.337 0.026 4.699 0.000 0.924 2.246   
## Complication\_riskLow 1.259 0.363 0.020 3.468 0.001 0.547 1.970   
## Complication\_riskHigh 0.327 0.315 0.006 1.039 0.299 -0.290 0.943   
## ArthritisYes 0.665 0.288 0.012 2.309 0.021 0.101 1.229   
## AnxietyYes 0.567 0.295 0.010 1.921 0.055 -0.011 1.146   
## Allergic\_rhinitisYes 0.401 0.282 0.007 1.421 0.155 -0.152 0.955   
## Reflux\_esophagitisYes 0.416 0.280 0.008 1.483 0.138 -0.134 0.965   
## ServicesCT Scan -1.057 0.438 -0.013 -2.415 0.016 -1.915 -0.199   
## ServicesIntravenous -0.095 0.311 -0.002 -0.304 0.761 -0.705 0.516   
## ServicesMRI -0.144 0.733 -0.001 -0.197 0.844 -1.580 1.292   
## Additional\_charges 0.000 0.000 -0.016 -2.106 0.035 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Allergic\_rhinitis   
##   
## Backward Elimination: Step 18   
##   
## Variable Allergic\_rhinitis Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.787   
## R-Squared 0.726 Coef. Var 40.015   
## Adj. R-Squared 0.725 MSE 190.093   
## Pred R-Squared 0.725 MAE 10.211   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5022852.672 13 386373.282 2032.548 0.0000   
## Residual 1898269.549 9986 190.093   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.736 0.499 31.522 0.000 14.758 16.715   
## Age 0.018 0.010 0.014 1.862 0.063 -0.001 0.037   
## ReAdmisYes 46.482 0.286 0.852 162.351 0.000 45.920 47.043   
## Initial\_adminObservation Admission 1.321 0.341 0.022 3.878 0.000 0.653 1.988   
## Initial\_adminElective Admission 1.590 0.337 0.026 4.714 0.000 0.929 2.251   
## Complication\_riskLow 1.268 0.363 0.020 3.493 0.000 0.556 1.979   
## Complication\_riskHigh 0.332 0.315 0.006 1.056 0.291 -0.285 0.949   
## ArthritisYes 0.668 0.288 0.012 2.322 0.020 0.104 1.233   
## AnxietyYes 0.569 0.295 0.010 1.927 0.054 -0.010 1.148   
## Reflux\_esophagitisYes 0.412 0.280 0.008 1.472 0.141 -0.137 0.962   
## ServicesCT Scan -1.053 0.438 -0.013 -2.406 0.016 -1.911 -0.195   
## ServicesIntravenous -0.096 0.311 -0.002 -0.307 0.759 -0.706 0.515   
## ServicesMRI -0.144 0.733 -0.001 -0.196 0.844 -1.580 1.292   
## Additional\_charges 0.000 0.000 -0.016 -2.091 0.037 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Reflux\_esophagitis   
##   
## Backward Elimination: Step 19   
##   
## Variable Reflux\_esophagitis Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.788   
## R-Squared 0.726 Coef. Var 40.018   
## Adj. R-Squared 0.725 MSE 190.115   
## Pred R-Squared 0.725 MAE 10.215   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5022440.840 12 418536.737 2201.489 0.0000   
## Residual 1898681.381 9987 190.115   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.915 0.484 32.868 0.000 14.966 16.864   
## Age 0.018 0.010 0.014 1.838 0.066 -0.001 0.036   
## ReAdmisYes 46.484 0.286 0.852 162.351 0.000 45.922 47.045   
## Initial\_adminObservation Admission 1.318 0.341 0.022 3.870 0.000 0.650 1.986   
## Initial\_adminElective Admission 1.593 0.337 0.026 4.723 0.000 0.932 2.255   
## Complication\_riskLow 1.270 0.363 0.020 3.497 0.000 0.558 1.981   
## Complication\_riskHigh 0.335 0.315 0.006 1.065 0.287 -0.282 0.952   
## ArthritisYes 0.675 0.288 0.012 2.345 0.019 0.111 1.239   
## AnxietyYes 0.566 0.295 0.010 1.916 0.055 -0.013 1.145   
## ServicesCT Scan -1.045 0.438 -0.013 -2.388 0.017 -1.903 -0.187   
## ServicesIntravenous -0.104 0.311 -0.002 -0.333 0.739 -0.714 0.507   
## ServicesMRI -0.143 0.733 -0.001 -0.195 0.845 -1.579 1.293   
## Additional\_charges 0.000 0.000 -0.016 -2.086 0.037 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
## - Services   
##   
## Backward Elimination: Step 20   
##   
## Variable Services Removed   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.790   
## R-Squared 0.726 Coef. Var 40.023   
## Adj. R-Squared 0.725 MSE 190.169   
## Pred R-Squared 0.725 MAE 10.219   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5021337.817 9 557926.424 2933.851 0.0000   
## Residual 1899784.404 9990 190.169   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.759 0.467 33.725 0.000 14.843 16.675   
## Age 0.018 0.010 0.014 1.833 0.067 -0.001 0.036   
## ReAdmisYes 46.468 0.286 0.851 162.344 0.000 45.907 47.029   
## Initial\_adminObservation Admission 1.331 0.341 0.022 3.909 0.000 0.664 1.999   
## Initial\_adminElective Admission 1.590 0.337 0.026 4.712 0.000 0.928 2.251   
## Complication\_riskLow 1.280 0.363 0.020 3.527 0.000 0.569 1.992   
## Complication\_riskHigh 0.332 0.315 0.006 1.056 0.291 -0.284 0.949   
## ArthritisYes 0.675 0.288 0.012 2.344 0.019 0.110 1.239   
## AnxietyYes 0.570 0.295 0.010 1.930 0.054 -0.009 1.149   
## Additional\_charges 0.000 0.000 -0.016 -2.104 0.035 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables satisfy the condition of p value = 0.1  
##   
##   
## Variables Removed:   
##   
## - Diabetes   
## - Gender   
## - Asthma   
## - Stroke   
## - Area   
## - Soft\_drink   
## - Children   
## - Overweight   
## - Income   
## - Marital   
## - BackPain   
## - VitD\_levels   
## - vitD\_supp   
## - HighBlood   
## - Doc\_visits   
## - Full\_meals\_eaten   
## - Hyperlipidemia   
## - Allergic\_rhinitis   
## - Reflux\_esophagitis   
## - Services   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.790   
## R-Squared 0.726 Coef. Var 40.023   
## Adj. R-Squared 0.725 MSE 190.169   
## Pred R-Squared 0.725 MAE 10.219   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5021337.817 9 557926.424 2933.851 0.0000   
## Residual 1899784.404 9990 190.169   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.759 0.467 33.725 0.000 14.843 16.675   
## Age 0.018 0.010 0.014 1.833 0.067 -0.001 0.036   
## ReAdmisYes 46.468 0.286 0.851 162.344 0.000 45.907 47.029   
## Initial\_adminObservation Admission 1.331 0.341 0.022 3.909 0.000 0.664 1.999   
## Initial\_adminElective Admission 1.590 0.337 0.026 4.712 0.000 0.928 2.251   
## Complication\_riskLow 1.280 0.363 0.020 3.527 0.000 0.569 1.992   
## Complication\_riskHigh 0.332 0.315 0.006 1.056 0.291 -0.284 0.949   
## ArthritisYes 0.675 0.288 0.012 2.344 0.019 0.110 1.239   
## AnxietyYes 0.570 0.295 0.010 1.930 0.054 -0.009 1.149   
## Additional\_charges 0.000 0.000 -0.016 -2.104 0.035 0.000 0.000   
## ----------------------------------------------------------------------------------------------------------------

# Fit forward selection stepwise regression model to mdl\_fwd   
mdl\_fwd <- ols\_step\_forward\_p(mdl\_fit, details = TRUE, penter = .1)

## Forward Selection Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1. Area   
## 2. Children   
## 3. Age   
## 4. Income   
## 5. Marital   
## 6. Gender   
## 7. ReAdmis   
## 8. VitD\_levels   
## 9. Doc\_visits   
## 10. Full\_meals\_eaten   
## 11. vitD\_supp   
## 12. Soft\_drink   
## 13. Initial\_admin   
## 14. HighBlood   
## 15. Stroke   
## 16. Complication\_risk   
## 17. Overweight   
## 18. Arthritis   
## 19. Diabetes   
## 20. Hyperlipidemia   
## 21. BackPain   
## 22. Anxiety   
## 23. Allergic\_rhinitis   
## 24. Reflux\_esophagitis   
## 25. Asthma   
## 26. Services   
## 27. Additional\_charges   
##   
## We are selecting variables based on p value...  
##   
##   
## Forward Selection: Step 1   
##   
## - ReAdmis   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.851 RMSE 13.823   
## R-Squared 0.724 Coef. Var 40.120   
## Adj. R-Squared 0.724 MSE 191.085   
## Pred R-Squared 0.724 MAE 10.233   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ----------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ----------------------------------------------------------------------------  
## Regression 5010653.468 1 5010653.468 26222.106 0.0000   
## Residual 1910468.753 9998 191.085   
## Total 6921122.221 9999   
## ----------------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 17.415 0.174 100.239 0.000 17.074 17.755   
## ReAdmisYes 46.445 0.287 0.851 161.932 0.000 45.883 47.007   
## -----------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 2   
##   
## - Initial\_admin   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.851 RMSE 13.805   
## R-Squared 0.725 Coef. Var 40.065   
## Adj. R-Squared 0.725 MSE 190.567   
## Pred R-Squared 0.725 MAE 10.224   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5016212.447 3 1672070.816 8774.179 0.0000   
## Residual 1904909.774 9996 190.567   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.673 0.222 75.103 0.000 16.237 17.108   
## ReAdmisYes 46.475 0.286 0.851 162.226 0.000 45.913 47.036   
## Initial\_adminObservation Admission 1.347 0.340 0.022 3.955 0.000 0.679 2.014   
## Initial\_adminElective Admission 1.610 0.337 0.027 4.772 0.000 0.949 2.271   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 3   
##   
## - Complication\_risk   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.797   
## R-Squared 0.725 Coef. Var 40.043   
## Adj. R-Squared 0.725 MSE 190.357   
## Pred R-Squared 0.725 MAE 10.222   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5018692.350 5 1003738.470 5272.921 0.0000   
## Residual 1902429.871 9994 190.357   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.293 0.270 60.354 0.000 15.764 16.822   
## ReAdmisYes 46.475 0.286 0.851 162.314 0.000 45.914 47.036   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.619 0.337 0.027 4.801 0.000 0.958 2.280   
## Complication\_riskLow 1.301 0.363 0.020 3.584 0.000 0.589 2.012   
## Complication\_riskHigh 0.291 0.314 0.005 0.927 0.354 -0.325 0.908   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 4   
##   
## - Arthritis   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.794   
## R-Squared 0.725 Coef. Var 40.034   
## Adj. R-Squared 0.725 MSE 190.269   
## Pred R-Squared 0.725 MAE 10.220   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5019760.051 6 836626.675 4397.064 0.0000   
## Residual 1901362.170 9993 190.269   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.045 0.289 55.428 0.000 15.478 16.612   
## ReAdmisYes 46.470 0.286 0.851 162.329 0.000 45.909 47.031   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.618 0.337 0.027 4.800 0.000 0.957 2.279   
## Complication\_riskLow 1.305 0.363 0.020 3.597 0.000 0.594 2.017   
## Complication\_riskHigh 0.307 0.314 0.006 0.977 0.329 -0.309 0.924   
## ArthritisYes 0.682 0.288 0.012 2.369 0.018 0.118 1.246   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 5   
##   
## - Anxiety   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.792   
## R-Squared 0.725 Coef. Var 40.029   
## Adj. R-Squared 0.725 MSE 190.219   
## Pred R-Squared 0.725 MAE 10.218   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5020455.112 7 717207.873 3770.435 0.0000   
## Residual 1900667.109 9992 190.219   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.863 0.305 52.053 0.000 15.265 16.460   
## ReAdmisYes 46.469 0.286 0.851 162.346 0.000 45.908 47.030   
## Initial\_adminObservation Admission 1.363 0.340 0.022 4.006 0.000 0.696 2.030   
## Initial\_adminElective Admission 1.623 0.337 0.027 4.815 0.000 0.962 2.284   
## Complication\_riskLow 1.308 0.363 0.020 3.604 0.000 0.597 2.019   
## Complication\_riskHigh 0.310 0.314 0.006 0.985 0.325 -0.307 0.926   
## ArthritisYes 0.675 0.288 0.012 2.346 0.019 0.111 1.240   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables to be added.  
##   
## Variables Entered:   
##   
## + ReAdmis   
## + Initial\_admin   
## + Complication\_risk   
## + Arthritis   
## + Anxiety   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.792   
## R-Squared 0.725 Coef. Var 40.029   
## Adj. R-Squared 0.725 MSE 190.219   
## Pred R-Squared 0.725 MAE 10.218   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5020455.112 7 717207.873 3770.435 0.0000   
## Residual 1900667.109 9992 190.219   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.863 0.305 52.053 0.000 15.265 16.460   
## ReAdmisYes 46.469 0.286 0.851 162.346 0.000 45.908 47.030   
## Initial\_adminObservation Admission 1.363 0.340 0.022 4.006 0.000 0.696 2.030   
## Initial\_adminElective Admission 1.623 0.337 0.027 4.815 0.000 0.962 2.284   
## Complication\_riskLow 1.308 0.363 0.020 3.604 0.000 0.597 2.019   
## Complication\_riskHigh 0.310 0.314 0.006 0.985 0.325 -0.307 0.926   
## ArthritisYes 0.675 0.288 0.012 2.346 0.019 0.111 1.240   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## ----------------------------------------------------------------------------------------------------------------

# Fit bidirectional stepwise regression model to mdl\_both   
mdl\_both <- ols\_step\_both\_p(mdl\_fit, details = TRUE, pent = .1, prem = .1)

## Stepwise Selection Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1. Area   
## 2. Children   
## 3. Age   
## 4. Income   
## 5. Marital   
## 6. Gender   
## 7. ReAdmis   
## 8. VitD\_levels   
## 9. Doc\_visits   
## 10. Full\_meals\_eaten   
## 11. vitD\_supp   
## 12. Soft\_drink   
## 13. Initial\_admin   
## 14. HighBlood   
## 15. Stroke   
## 16. Complication\_risk   
## 17. Overweight   
## 18. Arthritis   
## 19. Diabetes   
## 20. Hyperlipidemia   
## 21. BackPain   
## 22. Anxiety   
## 23. Allergic\_rhinitis   
## 24. Reflux\_esophagitis   
## 25. Asthma   
## 26. Services   
## 27. Additional\_charges   
##   
## We are selecting variables based on p value...  
##   
##   
## Stepwise Selection: Step 1   
##   
## - ReAdmis added   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.851 RMSE 13.823   
## R-Squared 0.724 Coef. Var 40.120   
## Adj. R-Squared 0.724 MSE 191.085   
## Pred R-Squared 0.724 MAE 10.233   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ----------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ----------------------------------------------------------------------------  
## Regression 5010653.468 1 5010653.468 26222.106 0.0000   
## Residual 1910468.753 9998 191.085   
## Total 6921122.221 9999   
## ----------------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 17.415 0.174 100.239 0.000 17.074 17.755   
## ReAdmisYes 46.445 0.287 0.851 161.932 0.000 45.883 47.007   
## -----------------------------------------------------------------------------------------  
##   
##   
##   
## Stepwise Selection: Step 2   
##   
## - Initial\_admin added   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.851 RMSE 13.805   
## R-Squared 0.725 Coef. Var 40.065   
## Adj. R-Squared 0.725 MSE 190.567   
## Pred R-Squared 0.725 MAE 10.224   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5016212.447 3 1672070.816 8774.179 0.0000   
## Residual 1904909.774 9996 190.567   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.673 0.222 75.103 0.000 16.237 17.108   
## ReAdmisYes 46.475 0.286 0.851 162.226 0.000 45.913 47.036   
## Initial\_adminObservation Admission 1.347 0.340 0.022 3.955 0.000 0.679 2.014   
## Initial\_adminElective Admission 1.610 0.337 0.027 4.772 0.000 0.949 2.271   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.851 RMSE 13.805   
## R-Squared 0.725 Coef. Var 40.065   
## Adj. R-Squared 0.725 MSE 190.567   
## Pred R-Squared 0.725 MAE 10.224   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5016212.447 3 1672070.816 8774.179 0.0000   
## Residual 1904909.774 9996 190.567   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.673 0.222 75.103 0.000 16.237 17.108   
## ReAdmisYes 46.475 0.286 0.851 162.226 0.000 45.913 47.036   
## Initial\_adminObservation Admission 1.347 0.340 0.022 3.955 0.000 0.679 2.014   
## Initial\_adminElective Admission 1.610 0.337 0.027 4.772 0.000 0.949 2.271   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Stepwise Selection: Step 3   
##   
## - Complication\_risk added   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.797   
## R-Squared 0.725 Coef. Var 40.043   
## Adj. R-Squared 0.725 MSE 190.357   
## Pred R-Squared 0.725 MAE 10.222   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5018692.350 5 1003738.470 5272.921 0.0000   
## Residual 1902429.871 9994 190.357   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.293 0.270 60.354 0.000 15.764 16.822   
## ReAdmisYes 46.475 0.286 0.851 162.314 0.000 45.914 47.036   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.619 0.337 0.027 4.801 0.000 0.958 2.280   
## Complication\_riskLow 1.301 0.363 0.020 3.584 0.000 0.589 2.012   
## Complication\_riskHigh 0.291 0.314 0.005 0.927 0.354 -0.325 0.908   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.797   
## R-Squared 0.725 Coef. Var 40.043   
## Adj. R-Squared 0.725 MSE 190.357   
## Pred R-Squared 0.725 MAE 10.222   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5018692.350 5 1003738.470 5272.921 0.0000   
## Residual 1902429.871 9994 190.357   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.293 0.270 60.354 0.000 15.764 16.822   
## ReAdmisYes 46.475 0.286 0.851 162.314 0.000 45.914 47.036   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.619 0.337 0.027 4.801 0.000 0.958 2.280   
## Complication\_riskLow 1.301 0.363 0.020 3.584 0.000 0.589 2.012   
## Complication\_riskHigh 0.291 0.314 0.005 0.927 0.354 -0.325 0.908   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Stepwise Selection: Step 4   
##   
## - Arthritis added   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.794   
## R-Squared 0.725 Coef. Var 40.034   
## Adj. R-Squared 0.725 MSE 190.269   
## Pred R-Squared 0.725 MAE 10.220   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5019760.051 6 836626.675 4397.064 0.0000   
## Residual 1901362.170 9993 190.269   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.045 0.289 55.428 0.000 15.478 16.612   
## ReAdmisYes 46.470 0.286 0.851 162.329 0.000 45.909 47.031   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.618 0.337 0.027 4.800 0.000 0.957 2.279   
## Complication\_riskLow 1.305 0.363 0.020 3.597 0.000 0.594 2.017   
## Complication\_riskHigh 0.307 0.314 0.006 0.977 0.329 -0.309 0.924   
## ArthritisYes 0.682 0.288 0.012 2.369 0.018 0.118 1.246   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.794   
## R-Squared 0.725 Coef. Var 40.034   
## Adj. R-Squared 0.725 MSE 190.269   
## Pred R-Squared 0.725 MAE 10.220   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5019760.051 6 836626.675 4397.064 0.0000   
## Residual 1901362.170 9993 190.269   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 16.045 0.289 55.428 0.000 15.478 16.612   
## ReAdmisYes 46.470 0.286 0.851 162.329 0.000 45.909 47.031   
## Initial\_adminObservation Admission 1.359 0.340 0.022 3.993 0.000 0.692 2.026   
## Initial\_adminElective Admission 1.618 0.337 0.027 4.800 0.000 0.957 2.279   
## Complication\_riskLow 1.305 0.363 0.020 3.597 0.000 0.594 2.017   
## Complication\_riskHigh 0.307 0.314 0.006 0.977 0.329 -0.309 0.924   
## ArthritisYes 0.682 0.288 0.012 2.369 0.018 0.118 1.246   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Stepwise Selection: Step 5   
##   
## - Anxiety added   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.792   
## R-Squared 0.725 Coef. Var 40.029   
## Adj. R-Squared 0.725 MSE 190.219   
## Pred R-Squared 0.725 MAE 10.218   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5020455.112 7 717207.873 3770.435 0.0000   
## Residual 1900667.109 9992 190.219   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.863 0.305 52.053 0.000 15.265 16.460   
## ReAdmisYes 46.469 0.286 0.851 162.346 0.000 45.908 47.030   
## Initial\_adminObservation Admission 1.363 0.340 0.022 4.006 0.000 0.696 2.030   
## Initial\_adminElective Admission 1.623 0.337 0.027 4.815 0.000 0.962 2.284   
## Complication\_riskLow 1.308 0.363 0.020 3.604 0.000 0.597 2.019   
## Complication\_riskHigh 0.310 0.314 0.006 0.985 0.325 -0.307 0.926   
## ArthritisYes 0.675 0.288 0.012 2.346 0.019 0.111 1.240   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.792   
## R-Squared 0.725 Coef. Var 40.029   
## Adj. R-Squared 0.725 MSE 190.219   
## Pred R-Squared 0.725 MAE 10.218   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5020455.112 7 717207.873 3770.435 0.0000   
## Residual 1900667.109 9992 190.219   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.863 0.305 52.053 0.000 15.265 16.460   
## ReAdmisYes 46.469 0.286 0.851 162.346 0.000 45.908 47.030   
## Initial\_adminObservation Admission 1.363 0.340 0.022 4.006 0.000 0.696 2.030   
## Initial\_adminElective Admission 1.623 0.337 0.027 4.815 0.000 0.962 2.284   
## Complication\_riskLow 1.308 0.363 0.020 3.604 0.000 0.597 2.019   
## Complication\_riskHigh 0.310 0.314 0.006 0.985 0.325 -0.307 0.926   
## ArthritisYes 0.675 0.288 0.012 2.346 0.019 0.111 1.240   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## ----------------------------------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables to be added/removed.  
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## ---------------------------------------------------------------  
## R 0.852 RMSE 13.792   
## R-Squared 0.725 Coef. Var 40.029   
## Adj. R-Squared 0.725 MSE 190.219   
## Pred R-Squared 0.725 MAE 10.218   
## ---------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## ---------------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## ---------------------------------------------------------------------------  
## Regression 5020455.112 7 717207.873 3770.435 0.0000   
## Residual 1900667.109 9992 190.219   
## Total 6921122.221 9999   
## ---------------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------------------------------  
## (Intercept) 15.863 0.305 52.053 0.000 15.265 16.460   
## ReAdmisYes 46.469 0.286 0.851 162.346 0.000 45.908 47.030   
## Initial\_adminObservation Admission 1.363 0.340 0.022 4.006 0.000 0.696 2.030   
## Initial\_adminElective Admission 1.623 0.337 0.027 4.815 0.000 0.962 2.284   
## Complication\_riskLow 1.308 0.363 0.020 3.604 0.000 0.597 2.019   
## Complication\_riskHigh 0.310 0.314 0.006 0.985 0.325 -0.307 0.926   
## ArthritisYes 0.675 0.288 0.012 2.346 0.019 0.111 1.240   
## AnxietyYes 0.565 0.295 0.010 1.912 0.056 -0.014 1.143   
## ----------------------------------------------------------------------------------------------------------------

# Summary of the backward model  
summary(mdl\_back$model)

##   
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),   
## data = l)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.296 -9.671 -3.406 4.212 41.812   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.576e+01 4.673e-01 33.725 < 2e-16 \*\*\*  
## Age 1.760e-02 9.603e-03 1.833 0.066799 .   
## ReAdmisYes 4.647e+01 2.862e-01 162.344 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 1.331e+00 3.406e-01 3.909 9.32e-05 \*\*\*  
## Initial\_adminElective Admission 1.590e+00 3.374e-01 4.712 2.48e-06 \*\*\*  
## Complication\_riskLow 1.280e+00 3.630e-01 3.527 0.000423 \*\*\*  
## Complication\_riskHigh 3.323e-01 3.146e-01 1.056 0.290886   
## ArthritisYes 6.747e-01 2.879e-01 2.344 0.019103 \*   
## AnxietyYes 5.699e-01 2.953e-01 1.930 0.053646 .   
## Additional\_charges -6.383e-05 3.034e-05 -2.104 0.035401 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9990 degrees of freedom  
## Multiple R-squared: 0.7255, Adjusted R-squared: 0.7253   
## F-statistic: 2934 on 9 and 9990 DF, p-value: < 2.2e-16

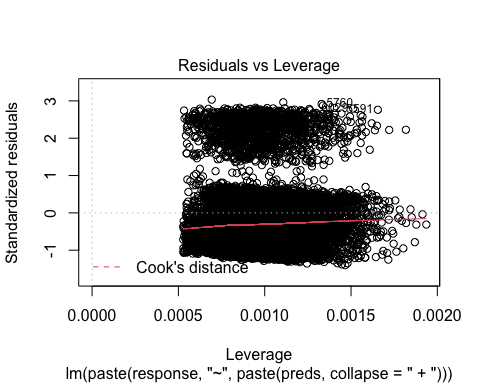
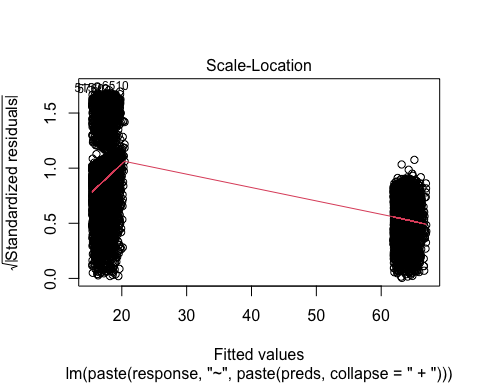
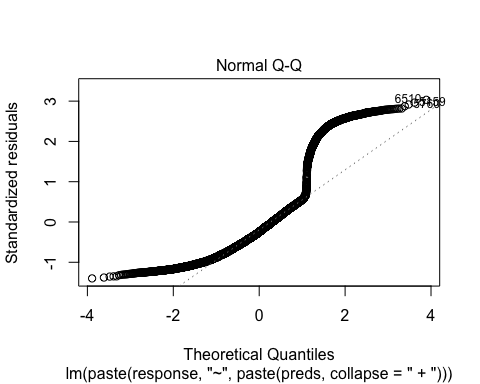
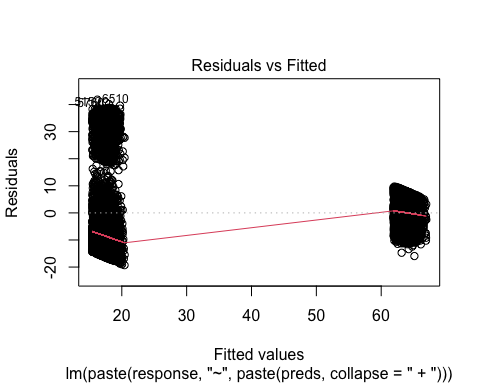
# Summary of the forward model  
summary(mdl\_fwd$model)

##   
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),   
## data = l)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.000 -9.647 -3.375 4.218 42.178   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.8628 0.3047 52.053 < 2e-16 \*\*\*  
## ReAdmisYes 46.4686 0.2862 162.346 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 1.3631 0.3402 4.006 6.22e-05 \*\*\*  
## Initial\_adminElective Admission 1.6229 0.3371 4.815 1.50e-06 \*\*\*  
## Complication\_riskLow 1.3078 0.3628 3.604 0.000314 \*\*\*  
## Complication\_riskHigh 0.3096 0.3145 0.985 0.324843   
## ArthritisYes 0.6754 0.2879 2.346 0.018988 \*   
## AnxietyYes 0.5645 0.2953 1.912 0.055963 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9992 degrees of freedom  
## Multiple R-squared: 0.7254, Adjusted R-squared: 0.7252   
## F-statistic: 3770 on 7 and 9992 DF, p-value: < 2.2e-16

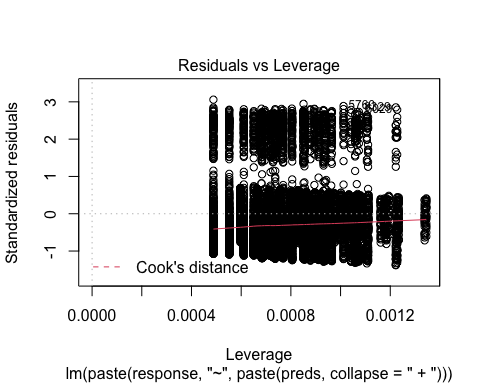
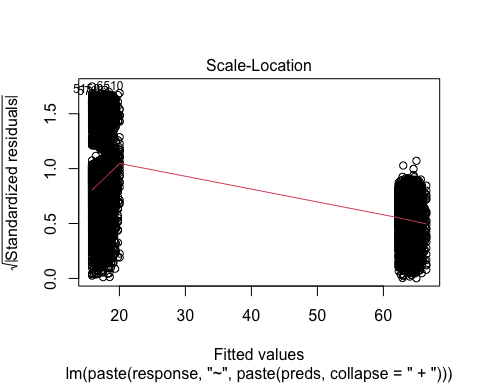
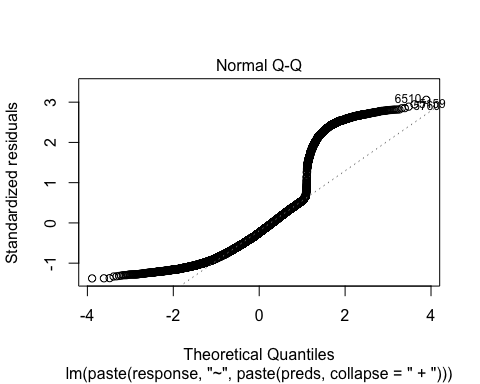
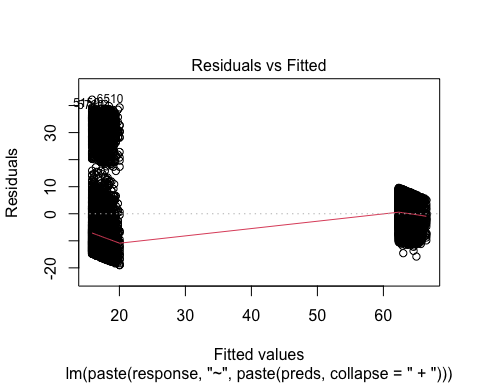
# Summary of the bidirectional model  
summary(mdl\_both$model)

##   
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),   
## data = l)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.000 -9.647 -3.375 4.218 42.178   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.8628 0.3047 52.053 < 2e-16 \*\*\*  
## ReAdmisYes 46.4686 0.2862 162.346 < 2e-16 \*\*\*  
## Initial\_adminObservation Admission 1.3631 0.3402 4.006 6.22e-05 \*\*\*  
## Initial\_adminElective Admission 1.6229 0.3371 4.815 1.50e-06 \*\*\*  
## Complication\_riskLow 1.3078 0.3628 3.604 0.000314 \*\*\*  
## Complication\_riskHigh 0.3096 0.3145 0.985 0.324843   
## ArthritisYes 0.6754 0.2879 2.346 0.018988 \*   
## AnxietyYes 0.5645 0.2953 1.912 0.055963 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.79 on 9992 degrees of freedom  
## Multiple R-squared: 0.7254, Adjusted R-squared: 0.7252   
## F-statistic: 3770 on 7 and 9992 DF, p-value: < 2.2e-16

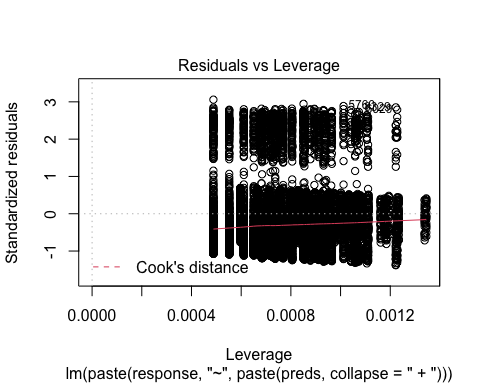
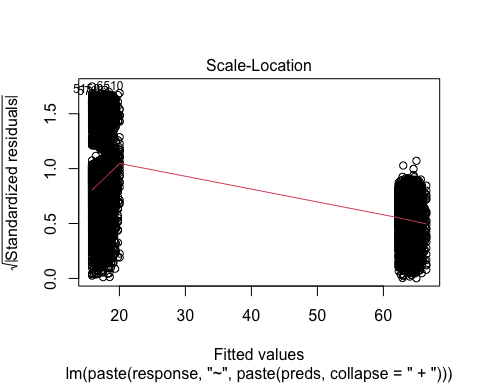
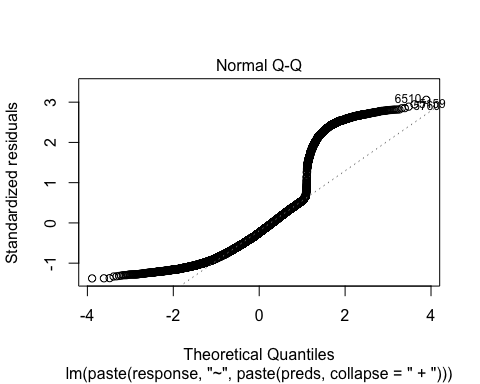
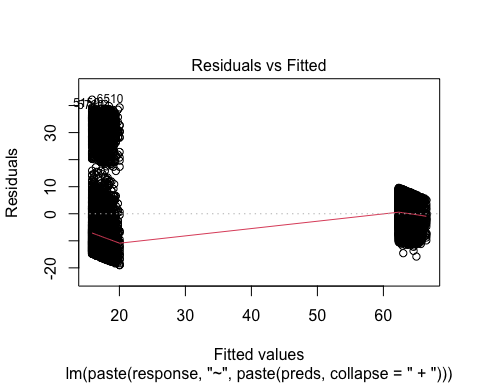
plot(mdl\_back$model)



plot(mdl\_fwd$model)



plot(mdl\_both$model)



## Part VI

### F1: Results

Multiple linear regression analyses were performed to interrogate the relationship between the response variable, Initial\_days, and a group of various, potentially related predictor variables as the initial model. Subsequently, three variable selection methods were employed, namely backward elimination, forward selection, and bidirectional stepwise, to iteratively reduce the initial model. The regression equation for the reduced model is:

A potential interpretation of this model and its coefficients could be:

Statistically, this model accounts for a reasonable amount of explained variance. Considering the data at our disposal, it’s perhaps the best approximation of estimating initial length of hospitalization using multiple regression available. Additionally, the computed F-statistic p-value for the model indicates a high level of significance. On the other hand, the practical significance of this model is of questionable value. One would need to know a given patient’s readmission status and amount billed for additional charges in order to have the necessary inputs to the model.

Perhaps most glaringly, two of the variables the model utilizes, ReAdmis and Additional\_charges, are information (or data) that cannot be known prior to hospitalization. Thus, the exercise of attempting to estimate or predict a patient’s length of stay prior to, or even to some extent during, admission to the hospital using this model is effectively impossible. Therefore, use of this model for analysis, while statistically significant, is likely only possible retrospectively and not as a prospective tool.

Additionally, the ReAdmis variable is of such critical relevance to the model as to render the rest of the variables in the model essentially moot. With a coefficient of days, it is clearly the primary constituent of the model. There is, however, value in understanding the impact of the other variables in relation to readmission status as well as to each other.

All of that aside, perhaps a more prescient and salient inquiry would be to investigate the quality of our data; specifically with regard to the composition of the Initial\_days variable. Given that the distribution of the variable is skewed and resembles a bimodal or “u-shaped” distribution, finding a linear relationship with the other variables (other than TotalCharge and Additional\_charges which seem to evidently be derived from Initial\_days anyway) makes multiple regression especially challenging and ineffective.

### F2: Recommendations

As a result of the preceding analysis, it is recommended that leadership consider implementing policy and procedure utilizing this multiple regression model to estimate a patient’s potential initial length of hospitalization. This model can be used as a limited, additional tool to estimate, and attempt to mitigate, longer hospitalizations which could potentially 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