

## **I. Motivation**

In 2016, total health expenditures in the United States exceeded \$3 trillion (“National Health Expenditures 2016 Highlights”). Over a quarter of these expenses—32 percent —were related to hospital care (“National Health Expenditures 2016 Highlights”). Health care spending in the U.S. is notoriously greater than in other countries of similar size and wealth; on average, our expenditures are double that of comparable countries (Cox). Despite health care being one of the most important sectors in the U.S., accounting for more than 17 percent of Gross Domestic Product, the widening of this gap demonstrates a need to keep spending under control (“National Health Expenditures 2016 Highlights”).

For the purpose of this project we adopted a medical and health services administration perspective. Although health care institutions are typically nonprofit—and are therefore not focused on return in a traditional business sense—they require capital to function and serve their communities. Additionally, capacity allocation is an important issue that hospitals face; they have limited resources (e.g., number of hospital beds) to provide to patients. We were interested in exploring methods to predict both the total cost of an inpatient stay as well as the length of the stay.

## **II. Description of Data**

Before conducting our analysis, some processing and cleaning of the data was needed. Please see the appendix for further detail. The original dataset can also be found on the New York State Department of Health’s website (“Hospital Inpatient Discharges (SPARCS De-Identified) Downloadable File: 2015”). The data used to conduct our analysis were 55,885 hospital inpatient discharge records from Mount Sinai Hospital in New York City, NY for 2015. A list of the variables included in the dataset used to build our models is included in the Appendix (Table 1). The variables which we expected to be most relevant were Type.of.Admission, APR.Severity.of.Illness, and APR.Risk.of.Mortality. The Length.of.Stay.Category variable (with factor levels “Up to 1 day”, “2 to 3 days”, “4 to 6 days”, and “7+ days”) was created using the Length.of.Stay variable and was introduced for classification purposes.

As we examined the data, it became clear that the Total.Costs and Length.of.Stay data was not normally distributed. Please see Appendix Figures 1 and 2 for a comparison of Total.Costs and the transformation of Total.Costs as well as Length.of.Stay and log.Length.of.Stay. We replaced the Total.Costs and Length.of.Stay variables with the log.Total.Costs and log.Length.of.Stay variables for analysis.

### III. Analysis

#### *Analysis 1 - Total Linear Model*

First we created a Linear Regression model predicting Total.Costs as a function of all 19 explanatory variables. Please see Appendix R Output 1. This model had an  $OSR^2$  of 77.96%. Such a high out-of-sample  $R^2$  is much higher than we would have expected for a model dealing with complex real-world data at this scale. This initial model was able to account for such a high percentage of variation in the data due to one of the explanatory variables, log.Length.of.Stay. This variable was one of the most significant for the model. However, it is also highly correlated with the response variable log.Total.Costs at 64.38%.

#### *Analysis 2 - Simple Linear Models*

Upon reflection, it did not make sense to include this variable in a regression model attempting to predict the cost of an inpatient stay; health services administrators would not have access to such information until after a patient had been discharged. Therefore, we removed the log.Length.of.Stay variable from further models attempting to predict log.Total.Costs. We then created 2 Linear Regression models, looking to explain log.Total.Costs in one and log.Length.of.Stay in the other. Each model excluded only the other dependant variable. Please see the Appendix R Output 2 and R Output 3.

The model predicting log.Total.Costs had an  $OSR^2$  of 62.89%; removing log.Length.of.Stay resulted in a reduction of approximately 15%. Although the  $OSR^2$  of this model is significantly less, it is more realistic for our dataset. It is able to give better cost predictions based on the data captured right when a patient is admitted to a hospital, as opposed to when they are ready to be discharged.

The model estimating log.Length.of.Stay had an  $OSR^2$  of 50.67%, and this used every independent variable in the data set as well. It appears that it is easier to predict total costs than length of stay when a patient first enters the hospital looking at these base models. To get more accurate predictions, several other methods were used, and are described below. Their effectiveness is compared against the base model to measure how “good” they are.

### *Analysis 3 - Regression Tree for Total Costs*

We next created a Regression Tree model predicting `log.Total.Costs` as a function of the 18 explanatory variables (again, excluding `log.Length.of.Stay`). We experimented with various values of `minbucket` (i.e., 50, 100, 1000, 2500), but eventually selected 500 observations, which yielded the best predictions when combined with the appropriate complexity parameter. Please see Appendix R Output 4 for relevant output. With such a large dataset and such a small relative `minbucket` value, we lost one of the main advantages of this method: a readable visual output.

With 29 terminal nodes, our tree was too large to allow us to see each individual split. However, we were able to approximate the splits at the top of our final tree using other `minbucket` and complexity parameter values. The initial split was “`Birth.Weight ≥ 2150`”. The subsequent node along the “yes” path is “`Race = Wht`”, resulting in two terminal nodes. If a patient’s birth weight is greater than or equal to 2150 grams (i.e., 4.74 lbs.) and the patient is white, their expected `log.Total.Costs` is 7.9; if the patient is not white, their expected `log.Total.Costs` is 8.2. These predicted values are equivalent to \$2,697.28 and \$3,640.95, respectively. The subsequent node along the right “no” path is “`APR.Severity.of.Illness = 1, 2, 3`” with all subsequent splits originating from that node. If a patient does not have a birth weight greater than or equal to 2150 grams (which should, in this dataset, indicate a patient that did not given birth) and has an `APR.Severity.of.Illness` of 4 (i.e., Extreme), their expected `log.Total.Cost` is 11. This is equivalent to an inpatient stay cost of \$59,875.14.

The  $OSR^2$  of the regression tree was 63.10%, an improvement over the Linear Regression model of only 0.21%. This is a disappointing, but not entirely unexpected, result due to the lack of difference or diversity between the Linear Regression and Regression Tree models.

### *Analysis 4 - Classification Tree for Length of Stay Category*

In order to better predict `Length.of.Stay`, we created a Classification Tree model. Please see Appendix R Output 5 for additional detail. Our first step was to convert the `Length.of.Stay` variable from numeric to factor as described above in this report. Once the grouping factors were determined, we moved forward with a classification tree with the adjusted `Length.of.Stay` categorical variable. The resulting model contained seven terminal nodes and delivered an accuracy of 57.86%, resulting in significantly better predictions than the simple linear regression model (which yielded an  $OSR^2$  of 50.67%).

The initial split of the tree was “`APR.MDC.Code = 14, 15`” representing Pregnancy and Perinatal Period, respectively. The subsequent node along the “yes” path is the terminal node predicting an inpatient stay between 2 and 3 days. For the “no” path, the next split was determined by “`APR.Severity.of.Illness = 1, 2`” (i.e., Asymptomatic or Moderate). If the patient was “yes”,

asymptomatic or minimally ill, the splits continued. However, if the patient was “no”, meaning they were majorly or extremely ill, then they were classified in a terminal node of 7+ days.

From this model, we were able to conclude that pregnancy is a strong predictor of the length of patient’s hospital stay, in addition to their severity of the illness categorization. Additionally, predicting patients with 2-3 day and 7+ day stays is much easier than the other groups. A confusion matrix confirms this. The model correctly predicted 77.77% of 2-3 day patients and 76.1% of 7+ day patient stays, compared to 31.56% of 0-1 day patient stays and 13.05% of 4-6 day patient stays.

#### *Analysis 5 - Random Forest for Total Costs*

To find a better model, a Random Forest was built to predict how much inpatient stays would cost. Please see Appendix R Output 6 for further details. The model was built using mtry of seven as the recommended mtry was seven or fourteen. Since fourteen would have meant almost all variables were utilized, seven was chosen as a more useful number providing greater diversity to the model.

The most important variables to predict total costs according to this model were Birth.Weight, APR.Severity.of.Illness.Code, Type.of.Admission, and APR.MDC.Code. All of these can be determined early in a patient’s stay (or shortly following the birth of a child), allowing the hospital to get a quick prediction of costs for a specific patient. Additional variables with high importance include a APR.Medical.Surgical.Description, Patient.Disposition, APR.Risk.of.Mortality, and CCS.Procedure.Code.

This Random Forest model had an  $OSR^2$  of 68.92%, a significant increase from the linear model and regression tree (about 9.5%). While this improvement is good, hospitals need to consider the tradeoff of greater accuracy with the increased complexity and higher difficulty in interpreting the results. Smaller hospitals may not have the personnel, expertise, or management necessary to understand and use this model effectively.

#### *Analysis 6 - Random Forest for Length of Stay Category*

A Random Forest was built to compare to the classification tree above, attempting to improve our prediction for a patient’s length of stay. The mtry recommended was seventeen; however, the model was built using mtry ten for a more effective outcome using less variables in hopes of bolstering diversity.

From this model, the most significant variables for predicting length of stay were APR.Severity.of.Illness.Code, APR.MDC.Code, Patient.Disposition, CCS.Diagnosis.Code, and

CCS.Procedure.Code. These variables encompass data that can be collected when a patient first arrives at a hospital, increasing the likelihood of accurately predicting a patient's stay length. However, the model does not reveal which, for example, of the 25 MDC Codes are most important or why.

The Random Forest had an accuracy of 60.54%, a small increase over the classification tree (57.86%). Additionally, the information obtained from the random forest is not necessarily as insightful as that gathered from the classification tree. For example, it was clear from the classification tree that pregnancy helped predict a patient's length of stay. When looking at a confusion matrix, we see a slight increase in the ability to predict 0-1 day and 2-3 day stays. There was also a slight decrease in the prediction of 4-6 day and 7+ day stays.

#### *Analysis 7 - LDA for Length of Stay Category*

Our final analysis was Linear Discriminant Analysis for Length.of.Stay.Category. Since this LDA was taken on a factor that had more than 2 outputs, any tangible interpretation is difficult to decipher. However, we were still able to see how well the LDA split the dataset through its proportion of trace results. Proportion of trace measures how well the LDA splits the dataset with each cut or function it produces. Please see Appendix R Output 8 for additional detail.

The first function produced by LDA successfully split 76.58% of the data, with the following cut splitting 19.06% of the data and the final cut splitting the remaining 4.36% of the data. Since LDA is an exploratory test, it is difficult to say how effective this analysis would be on out of sample data, but it does provide useful information on variables that other hospitals similar to Mount Sinai can collect to improve their predictions for patient length of stay.

The highest impact variables (positive and negative) in the first function created by LDA are Patient.Disposition (which is not particularly useful), Payment.Typology., APR.Severity.of.Illness.Code, and APR.MDC.Code. While the information represented by these variables cannot always be collected when a patient arrives at a hospital, hospitals can use the information of past high cost patients with similar characteristics to predict costs in this case. Patients with a history of leaving medical care before being discharged by a doctor, for example, will likely be more expensive in the long run than other types of patients, as detailed by our Linear Discriminant Analysis output.

#### **IV. Summary of Results**

Health care administrators and hospital management can gain a few key takeaways from this analysis. The key predictors of length of stay and total cost were labor, severity of illness, and diagnostic MDC codes. Additionally, our models were most effective at predicting patients with 2-3 day (pregnancy/labor patients) and 7+ day stays (severity of illness of 3 or 4).

After testing a variety of models in our analysis, it is recommended that Mount Sinai Hospital use the Classification Tree model to derive insights from their inpatient data. Although the Random Forest offers marginally higher accuracy (+2.86%) than the Classification Tree, the Tree is more readily interpretable since management can take advantage of the visual output it provides. For future use, it is recommended that Mount Sinai share this model (and improvements of it) with smaller hospitals and health systems that can benefit from using this kind of analysis. Additionally, Mount Sinai can offer current best practice to these hospitals by stating the importance of quality over quantity in the variables tracked. As seen in the Linear Discriminant Analysis, there are only several high impact variables. Other hospitals should only use their limited resources to track key variables.

In conclusion, although this analysis will prove beneficial to Mount Sinai Hospital, there is always more analysis to be done. Further recommendations would be to break total inpatient stay cost into three buckets: low-cost, mid-cost, high-cost. Additionally, removing inpatient stays related to pregnancy or labor from the data may offer more insight into length of stay and total costs. And finally, tracking quality of care would be valuable in comparing cross-hospital data.

## Sources

Cox, Cynthia and Bradley Sawyer. “How does health spending in the U.S. compare to other countries?” *Peterson-Kaiser Health Systems Tracker*, 13 Feb 2018, <https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/#item-start>. Accessed 04 April 2018.

“Hospital Inpatient Discharges (SPARCS De-Identified) Downloadable File: 2015.” *Health Data NY*, 6 March 2018, <https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/4ep6-xmju>.

“Major Diagnostic Category.” *Wikipedia*, 4 Aug 2017, [https://en.wikipedia.org/w/index.php?title=Major\\_Diagnostic\\_Category](https://en.wikipedia.org/w/index.php?title=Major_Diagnostic_Category). Accessed 06 April 2018.

“National Health Expenditures 2016 Highlights.” *CMS.gov*, <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf>. Accessed 04 April 2018.

## Appendix

### Description of Processing:

We extracted a subset from “Hospital Inpatient Discharges (SPARCS De-Identified) Download File: 2015” for a single hospital in the state (New York State Department of Health). Mount Sinai Hospital (Permanent.Facility.Id: 1456) was the most-represented of the included facilities with 55,888 patient records. The approach used to separate Mount Sinai Hospital’s observations involved excluding abortion records (release of such information is limited due to HIPAA [Health Insurance Portability and Accountability Act of 1996] protections).

Once separated, we removed two observations that were missing explanatory variable information (i.e., Type.of.Admission of type “Not Available” and Patient.Disposition of type “Another Type Not Listed”). Additional formatting and conversions were also done (i.e., converting Total.Charges and Total.Costs from character to numeric values, replacement of “120 +” character response value for Length.of.Stay with 120 numeric response value). Factor level “Miscellaneous/Other” was substituted for a missing factor level of Payment.Typology.2 and Payment.Typology.3. Certain variables were excluded such as those containing hospital identification information and redundant variables (i.e., those that offered descriptions for an equivalent variable containing a categorization or code). Discharge.Year was also excluded since all records were for 2015.

Before building any models, we also explored the correlation amongst the explanatory variables. Although APR.DRG.Code and APR.MDC.Code are both factors, they represent similar information (“Major Diagnostic Category”). Diagnostic-related groups index to and provide more granular detail about the services provided to a patient than Major Diagnostic Categories (of which there are 25), so we chose to keep APR.MDC.Code in our dataset. High correlation between the numeric variables Length.of.Stay and Total.Charges prompted the removal of Total.Charges as well.

Lastly, we removed one outlier from our data: an observation which had a value of Total.Costs greater than \$2,000,000.00. This was an extreme outlier well exceeding the Total.Costs of other inpatient stays.



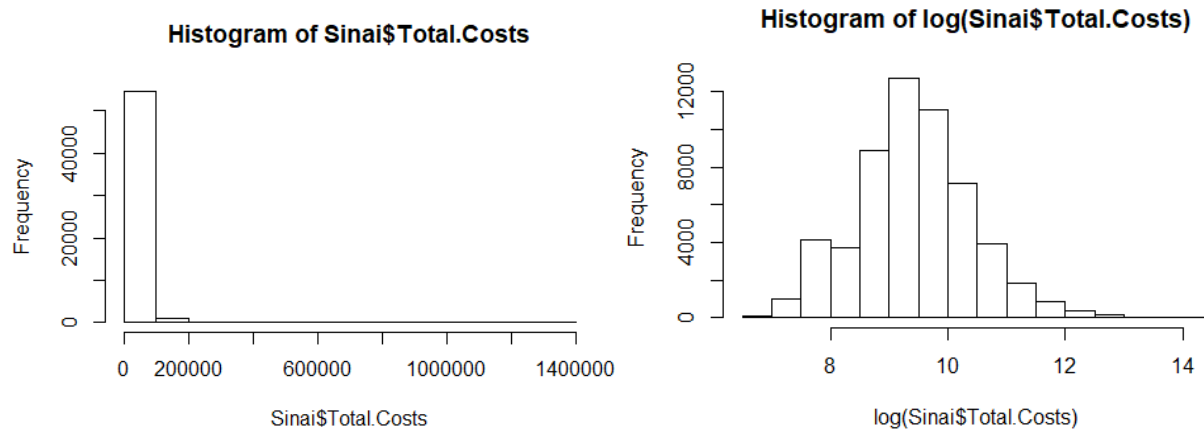
**Table 1:** Dataset Variable Names & Types

Variable Name	Variable Type
Age.Group	Factor with 5 levels: “0 to 17”, “18 to 29”, “30 to 49”, “50 to 69”, “70 or Older”
Gender	Factor with 2 levels: “F”, “M”
Race	Factor with 3 levels: “Black/African American”, “Other Race”, “White”
Ethnicity	Factor with 2 levels: “Not Span/Hispanic”, “Span/Hispanic”
Length.of.Stay	Integer
Type.of.Admission	Factor with 4 levels: “Elective”, “Emergency”, “Newborn”, “Urgent”
Patient.Disposition	Factor with 15 levels: “Cancer Center or Children’s Hospital”, “Court/Law Enforcement”, “Expired”, “Facility w/ Custodial/Supportive Care”, “Federal Health Care Facility”, “Home or Self Care”, “Home w/ Home Health Services”, “Hospice - Home”, “Hospice - Medical Facility”, “Inpatient Rehabilitation Facility”, “Left Against Medical Advice”, “Medicare Cert Long Term Care Hospital”, “Psychiatric Hospital or Unit of Hospital”, “Short-term Hospital”, “Skilled Nursing Home”
CCS.Diagnosis.Code	Integer
CCS.Procedure.Code	Integer
APR.MDC.Code	Factor with 25 levels: “1”, “2”, “3”, “4”, “5”, “6”, “7”, “8”, “9”, “10”, “11”, “12”, “13”, “14”, “15”, “16”, “17”, “18”, “19”, “20”, “21”, “22”, “23”, “24”, “25”
APR.Severity.of.Illness.Code	Factor with 4 levels: “1”, “2”, “3”, “4”
APR.Risk.of.Mortality	Factor with 4 levels: “Extreme”, “Major”, “Minor”, “Moderate”
APR.Medical.Surgical.Description	Factor with 2 levels: “Medical”, “Surgical”

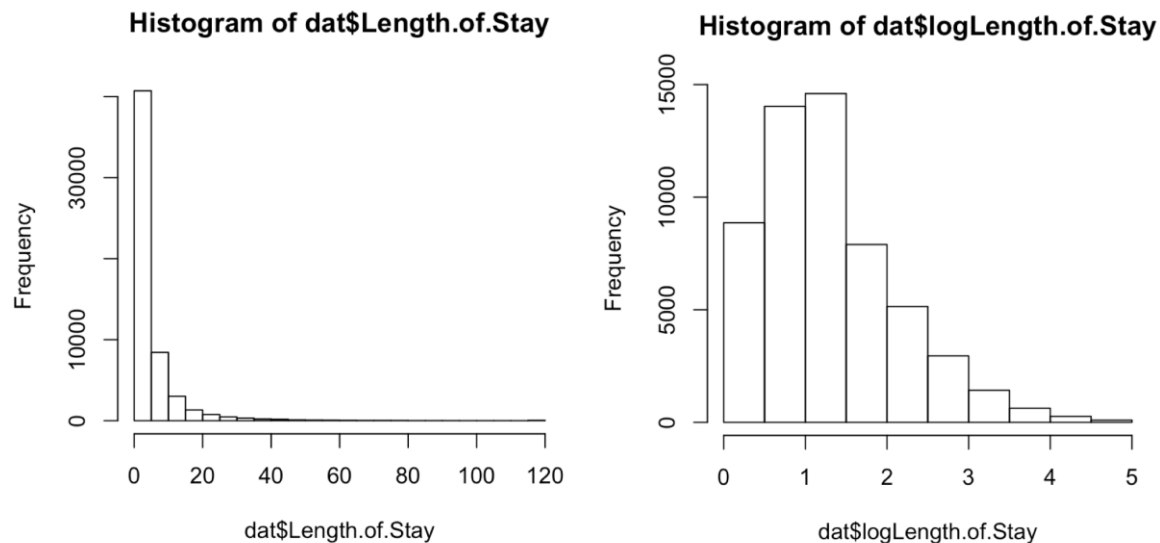
Payment.Typology.1	Factor with 8 levels: “Blue Cross/Blue Shield”, “Federal/State/Local/VA”, “Managed Care, Unspecified”, “Medicaid”, “Medicare”, “Miscellaneous/Other”, “Private Health Insurance”, “Self-Pay”
Payment.Typology.2	Factor with 8 levels: “Blue Cross/Blue Shield”, “Federal/State/Local/VA”, “Managed Care, Unspecified”, “Medicaid”, “Medicare”, “Miscellaneous/Other”, “Private Health Insurance”, “Self-Pay”
Payment.Typology.3	Factor with 8 levels: “Blue Cross/Blue Shield”, “Federal/State/Local/VA”, “Managed Care, Unspecified”, “Medicaid”, “Medicare”, “Miscellaneous/Other”, “Private Health Insurance”, “Self-Pay”
Birth.Weight	Integer
Emergency.Department.Indicator	Factor with 2 levels: “N”, “Y”
log.Total.Costs	Numeric (Logarithmic transformation of original numeric Total.Costs variable)
Length.of.Stay.Category	Factor with 4 levels: “Up to 1 day”, “2 to 3 days,” “4 to 6 days”, “7+ days”

**Figure 1:** Transformation of Total.Costs

```
> Sinai <- read.csv("Model.csv")
> hist(Sinai$Total.Costs)
> hist(log(Sinai$Total.Costs))
> view(Sinai)
> Sinai$log.Total.Costs <- log(Sinai$Total.Costs)
> Sinai <- Sinai[,-19]
```



**Figure 2:** Transformation of Length.of.Stay



**R Output 1:** Linear Regression with log.Length.of.Stay for log.Total.Costs

```
> set.seed(1760)
> split <- sample(nrow(Sinai), 0.75*nrow(Sinai))
> train <- Sinai[split,]
> test <- Sinai[-split,]
> model.lm <- lm(log.Total.Costs ~., data = train)
```

```
> summary(model.lin)
```

```
Call:
```

```
lm(formula = log.Total.Costs ~ ., data = train)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-3.9432 -0.2434  0.0045  0.2305  3.4097
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.115e+00	1.545e-01	58.994	< 2e-16
Age.Group18 to 29	-2.010e-01	1.318e-02	-15.244	< 2e-16
Age.Group30 to 49	-2.454e-01	1.201e-02	-20.432	< 2e-16
Age.Group50 to 69	-2.556e-01	1.174e-02	-21.774	< 2e-16
Age.Group70 or older	-3.094e-01	1.400e-02	-22.096	< 2e-16
GenderM	3.188e-02	5.106e-03	6.244	4.32e-10
RaceOther Race	7.454e-03	6.665e-03	1.118	0.26339
Racewhite	-1.566e-02	7.042e-03	-2.224	0.02617
EthnicitySpanish/Hispanic	1.275e-01	3.316e-01	0.384	0.70071
Length.of.Stay	5.411e-02	3.253e-04	166.359	< 2e-16
Type.of.AdmissionEmergency	-1.495e-01	3.420e-02	-4.371	1.24e-05
Type.of.AdmissionNewborn	-9.712e-01	4.198e-02	-23.132	< 2e-16
Type.of.AdmissionUrgent	-5.946e-02	8.770e-03	-6.781	1.21e-11
Patient.DispositionCourt/Law Enforcement	2.566e-01	2.232e-01	1.150	0.25030
Patient.DispositionExpired	3.291e-01	1.367e-01	2.408	0.01604
Patient.DispositionFacility w/ Custodial/Supportive Care	2.192e-01	1.469e-01	1.492	0.13570
Patient.DispositionFederal Health Care Facility	-2.367e-01	3.027e-01	-0.782	0.43435
Patient.DispositionHome or Self Care	1.040e-01	1.358e-01	0.766	0.44377
Patient.DispositionHome w/ Home Health Services	2.795e-01	1.358e-01	2.059	0.03954
Patient.DispositionHospice - Home	2.523e-01	1.395e-01	1.809	0.07051
Patient.DispositionHospice - Medical Facility	2.739e-01	1.389e-01	1.972	0.04865
Patient.DispositionInpatient Rehabilitation Facility	3.977e-01	1.372e-01	2.898	0.00376
Patient.DispositionLeft Against Medical Advice	-1.780e-01	1.388e-01	-1.283	0.19951
Patient.DispositionMedicare Cert Long Term Care Hospital	6.442e-02	1.435e-01	0.449	0.65346
Patient.DispositionPsychiatric Hospital or Unit of Hosp	4.122e-02	1.439e-01	0.286	0.77453
Patient.DispositionShort-term Hospital	4.376e-02	1.392e-01	0.314	0.75330
Patient.DispositionSkilled Nursing Home	2.892e-01	1.359e-01	2.128	0.03336
CCS.Diagnosis.Code	-9.202e-05	3.816e-05	-2.411	0.01590
CCS.Procedure.Code	5.225e-04	3.310e-05	15.785	< 2e-16
APR.MDC.Code2	-1.794e-01	7.308e-02	-2.455	0.01409
APR.MDC.Code3	-2.602e-01	2.184e-02	-11.911	< 2e-16
APR.MDC.Code4	-1.601e-01	1.462e-02	-10.949	< 2e-16
APR.MDC.Code5	-1.705e-02	1.190e-02	-1.434	0.15170
APR.MDC.Code6	-1.742e-01	1.294e-02	-13.461	< 2e-16
APR.MDC.Code7	-9.284e-02	1.657e-02	-5.601	2.14e-08
APR.MDC.Code8	6.017e-02	1.410e-02	4.268	1.97e-05
APR.MDC.Code9	-1.624e-01	2.003e-02	-8.110	5.20e-16
APR.MDC.Code10	-2.427e-01	1.756e-02	-13.819	< 2e-16
APR.MDC.Code11	-7.198e-02	1.594e-02	-4.517	6.30e-06
APR.MDC.Code12	-2.537e-01	2.620e-02	-9.683	< 2e-16
APR.MDC.Code13	-3.852e-01	2.157e-02	-17.863	< 2e-16
APR.MDC.Code14	-2.977e-01	1.437e-02	-20.709	< 2e-16
APR.MDC.Code15	2.217e-01	4.969e-02	4.461	8.20e-06
APR.MDC.Code16	7.000e-02	2.193e-02	3.192	0.00141
APR.MDC.Code17	1.253e-01	1.978e-02	6.334	2.41e-10
APR.MDC.Code18	-9.279e-02	1.583e-02	-5.860	4.65e-09
APR.MDC.Code19	7.024e-02	2.655e-02	2.646	0.00816
APR.MDC.Code20	-6.881e-02	4.878e-02	-1.411	0.15839
APR.MDC.Code21	-1.773e-01	2.708e-02	-6.548	5.89e-11
APR.MDC.Code22	1.577e-01	2.346e-01	0.672	0.50155
APR.MDC.Code23	-2.395e-01	2.014e-02	-11.888	< 2e-16
APR.MDC.Code24	-2.453e-02	3.570e-02	-0.687	0.49204
APR.MDC.Code25	-1.505e-01	1.662e-01	-0.906	0.36507
APR.Severity.of.Illness.Code2	1.592e-01	6.102e-03	26.089	< 2e-16
APR.Severity.of.Illness.Code3	3.744e-01	8.760e-03	42.734	< 2e-16
APR.Severity.of.Illness.Code4	5.308e-01	1.542e-02	34.428	< 2e-16
APR.Risk.of.MortalityMajor	2.490e-04	1.516e-02	0.016	0.98689
APR.Risk.of.MortalityMinor	-6.960e-02	1.767e-02	-3.939	8.18e-05
APR.Risk.of.MortalityModerate	-4.020e-02	1.653e-02	-2.433	0.01500
APR.Medical.Surgical.DescriptionSurgical	5.490e-01	6.578e-03	83.458	< 2e-16
Payment.Typeology.1Federal/State/Local/VA	-1.566e-01	9.063e-02	-1.728	0.08398
Payment.Typeology.1Managed Care, Unspecified	2.283e-02	2.371e-02	0.963	0.33553
Payment.Typeology.1Medicaid	4.844e-03	1.010e-02	0.479	0.63159
Payment.Typeology.1Medicare	-2.302e-02	1.159e-02	-1.986	0.04701
Payment.Typeology.1Miscellaneous/other	-5.688e-02	2.988e-02	-1.903	0.05699

Payment.Typeology.1Private Health Insurance	-5.332e-03	8.211e-03	-0.649	0.51613
Payment.Typeology.1Self-Pay	-1.762e-02	2.402e-02	-0.733	0.46329
Payment.Typeology.2Federal/State/Local/VA	3.430e-02	8.921e-02	0.384	0.70067
Payment.Typeology.2Managed Care, Unspecified	7.481e-02	5.068e-02	1.476	0.13992
Payment.Typeology.2Medicaid	-3.946e-02	1.258e-02	-3.137	0.00171
Payment.Typeology.2Medicare	-6.932e-03	1.295e-02	-0.535	0.59253
Payment.Typeology.2Miscellaneous/Other	-2.449e-01	1.103e-01	-2.220	0.02642
Payment.Typeology.2Private Health Insurance	-4.115e-03	1.301e-02	-0.316	0.75189
Payment.Typeology.2Self-Pay	-9.015e-02	9.553e-02	-0.944	0.34531
Payment.Typeology.3Federal/State/Local/VA	-1.350e-02	1.904e-01	-0.071	0.94349
Payment.Typeology.3Managed Care, Unspecified	1.233e-01	2.209e-01	0.558	0.57678
Payment.Typeology.3Medicaid	-1.736e-02	6.991e-02	-0.248	0.80384
Payment.Typeology.3Medicare	-4.265e-02	9.973e-02	-0.428	0.66891
Payment.Typeology.3Miscellaneous/Other	8.747e-02	1.176e-01	0.744	0.45683
Payment.Typeology.3Private Health Insurance	7.881e-02	8.284e-02	0.951	0.34141
Payment.Typeology.3Self-Pay	1.806e-02	6.959e-02	0.260	0.79520
Birth.Weight	-2.017e-04	8.200e-06	-24.603	< 2e-16
Emergency.Department.IndicatorY	6.020e-02	3.392e-02	1.775	0.07596
(Intercept)	***			
Age.Group18 to 29	***			
Age.Group30 to 49	***			
Age.Group50 to 69	***			
Age.Group70 or older	***			
GenderM	***			
RaceOther Race				
RaceWhite	*			
EthnicitySpanish/Hispanic				
Length.of.Stay	***			
Type.of.AdmissionEmergency	***			
Type.of.AdmissionNewborn	***			
Type.of.AdmissionUrgent	***			
Patient.DispositionCourt/Law Enforcement				
Patient.DispositionExpired	*			
Patient.DispositionFacility w/ Custodial/Supportive Care				
Patient.DispositionFederal Health Care Facility				
Patient.DispositionHome or Self Care				
Patient.DispositionHome w/ Home Health Services	*			
Patient.DispositionHospice - Home	.			
Patient.DispositionHospice - Medical Facility	*			
Patient.DispositionInpatient Rehabilitation Facility	**			
Patient.DispositionLeft Against Medical Advice				
Patient.DispositionMedicare Cert Long Term Care Hospital				
Patient.DispositionPsychiatric Hospital or Unit of Hosp				
Patient.DispositionShort-term Hospital				
Patient.DispositionSkilled Nursing Home	*			
CCS.Diagnosis.Code	*			
CCS.Procedure.Code	***			
APR.MDC.Code2	*			
APR.MDC.Code3	***			
APR.MDC.Code4	***			
APR.MDC.Code5				
APR.MDC.Code6	***			
APR.MDC.Code7	***			
APR.MDC.Code8	***			
APR.MDC.Code9	***			
APR.MDC.Code10	***			
APR.MDC.Code11	***			
APR.MDC.Code12	***			
APR.MDC.Code13	***			
APR.MDC.Code14	***			
APR.MDC.Code15	***			
APR.MDC.Code16	**			
APR.MDC.Code17	***			
APR.MDC.Code18	***			
APR.MDC.Code19	**			
APR.MDC.Code20				
APR.MDC.Code21	***			
APR.MDC.Code22				
APR.MDC.Code23	***			
APR.MDC.Code24				
APR.MDC.Code25				
APR.Severity.of.Illness.Code2	***			
APR.Severity.of.Illness.Code3	***			
APR.Severity.of.Illness.Code4	***			
APR.Risk.of.MortalityMajor	...			

```

APR.Risk.of.MortalityMinor          ***
APR.Risk.of.MortalityModerate       *
APR.Medical.Surgical.DescriptionSurgical ***
Payment.Typology.1Federal/State/Local/VA .
Payment.Typology.1Managed Care, Unspecified
Payment.Typology.1Medicaid
Payment.Typology.1Medicare          *
Payment.Typology.1Miscellaneous/Other .
Payment.Typology.1Private Health Insurance
Payment.Typology.1Self-Pay
Payment.Typology.2Federal/State/Local/VA
Payment.Typology.2Managed Care, Unspecified
Payment.Typology.2Medicaid          **
Payment.Typology.2Medicare
Payment.Typology.2Miscellaneous/Other *
Payment.Typology.2Private Health Insurance
Payment.Typology.2Self-Pay
Payment.Typology.3Federal/State/Local/VA
Payment.Typology.3Managed Care, Unspecified
Payment.Typology.3Medicaid
Payment.Typology.3Medicare
Payment.Typology.3Miscellaneous/Other
Payment.Typology.3Private Health Insurance
Payment.Typology.3Self-Pay
Birth.Weight                        ***
Emergency.Department.IndicatorY     .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4685 on 41830 degrees of freedom
Multiple R-squared:  0.7757,    Adjusted R-squared:  0.7753
F-statistic: 1764 on 82 and 41830 DF,  p-value: < 2.2e-16

> pred.lin <- predict(model.lin, newdata=test)
> SSE.lin <- sum((pred.lin - test$log.Total.Costs)^2)
> train.mean <- mean(train$log.Total.Costs)
> SST.lin <- sum((train.mean - test$log.Total.Costs)^2)
> OSR2.lin <- 1 - SSE.lin/SST.lin
> OSR2.lin
[1] 0.7795775

```

## R Output 2: Linear Regression without log.Length.of.Stay for log.Total.Costs

```

> datX <- sinai[,c(5,19)]
> cor(datX)
      Length.of.Stay log.Total.Costs
Length.of.Stay      1.0000000      0.6438248
log.Total.Costs      0.6438248      1.0000000
> sinai <- sinai[,-5]
> set.seed(1760)
> split <- sample(nrow(sinai), 0.75*nrow(sinai))
> train <- sinai[split,]
> test <- sinai[-split,]
> model.lin <- lm(log.Total.Costs ~., data = train)

```

```

> summary(model.lm)

Call:
lm(formula = log.Total.Costs ~ ., data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0832 -0.3412 -0.0231  0.2848  5.8065

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.894e+00  1.991e-01  49.699 < 2e-16
Age.Group18 to 29 -1.764e-01  1.699e-02 -10.379 < 2e-16
Age.Group30 to 49 -2.328e-01  1.548e-02 -15.041 < 2e-16
Age.Group50 to 69 -2.680e-01  1.513e-02 -17.710 < 2e-16
Age.Group70 or Older -4.006e-01  1.803e-02 -22.215 < 2e-16
GenderM        3.314e-02  6.581e-03   5.036 4.77e-07
RaceOther Race -7.211e-03  8.590e-03  -0.839 0.401254
RaceWhite     -2.765e-02  9.076e-03  -3.047 0.002315
EthnicitySpanish/Hispanic 6.459e-02  4.275e-01  0.151 0.879896
Type.of.AdmissionEmergency -1.826e-02  4.408e-02  -0.414 0.678596
Type.of.AdmissionNewborn -9.168e-01  5.412e-02 -16.941 < 2e-16
Type.of.AdmissionUrgent  5.574e-02  1.127e-02   4.946 7.60e-07
Patient.DispositionCourt/Law Enforcement 2.197e-02  2.877e-01  0.076 0.939111
Patient.DispositionExpired -2.785e-01  1.761e-01  -1.581 0.113832
Patient.DispositionFacility w/ Custodial/Supportive care -2.034e-01  1.894e-01  -1.074 0.282885
Patient.DispositionFederal Health Care Facility -1.034e+00  3.902e-01  -2.650 0.008048
Patient.DispositionHome or Self care -5.017e-01  1.750e-01  -2.867 0.004140
Patient.DispositionHome w/ Home Health Services -1.963e-01  1.750e-01  -1.122 0.261839
Patient.DispositionHospice - Home -3.165e-01  1.798e-01  -1.761 0.078264
Patient.DispositionHospice - Medical Facility -2.930e-01  1.790e-01  -1.636 0.101782
Patient.DispositionInpatient Rehabilitation Facility 6.715e-02  1.769e-01  0.380 0.704178
Patient.DispositionLeft Against Medical Advice -8.285e-01  1.788e-01  -4.634 3.61e-06
Patient.DispositionMedicare Cert Long Term Care Hospital -1.227e-01  1.849e-01  -0.664 0.506935
Patient.DispositionPsychiatric Hospital or Unit of Hosp -3.846e-01  1.855e-01  -2.074 0.038117
Patient.DispositionShort-term Hospital -3.471e-01  1.794e-01  -1.934 0.053101
Patient.DispositionSkilled Nursing Home 1.766e-02  1.752e-01  0.101 0.919726
CCS.Diagnosis.code 5.437e-05  4.918e-05  1.106 0.268897
CCS.Procedure.code 6.143e-04  4.266e-05  14.400 < 2e-16
APR.MDC.Code2 -2.316e-01  9.421e-02  -2.458 0.013967
APR.MDC.Code3 -3.004e-01  2.815e-02 -10.670 < 2e-16
APR.MDC.Code4 -1.930e-01  1.884e-02 -10.240 < 2e-16
APR.MDC.Code5 -1.059e-01  1.532e-02  -6.910 4.92e-12
APR.MDC.Code6 -1.612e-01  1.668e-02  -9.660 < 2e-16
APR.MDC.Code7 -1.039e-01  2.136e-02  -4.862 1.17e-06
APR.MDC.Code8 -5.489e-02  1.815e-02  -3.024 0.002494
APR.MDC.Code9 -1.994e-01  2.581e-02  -7.723 1.16e-14
APR.MDC.Code10 -3.104e-01  2.264e-02 -13.711 < 2e-16
APR.MDC.Code11 -1.178e-01  2.054e-02  -5.737 9.71e-09
APR.MDC.Code12 -3.109e-01  3.377e-02  -9.205 < 2e-16
APR.MDC.Code13 -4.336e-01  2.780e-02 -15.601 < 2e-16
APR.MDC.Code14 -2.780e-01  1.853e-02 -15.006 < 2e-16
APR.MDC.Code15 3.466e-01  6.405e-02  5.412 6.28e-08
APR.MDC.Code16 1.351e-01  2.827e-02  4.780 1.76e-06
APR.MDC.Code17 4.273e-01  2.539e-02  16.832 < 2e-16
APR.MDC.Code18 -1.571e-01  2.040e-02  -7.699 1.40e-14
APR.MDC.Code19 5.407e-01  3.403e-02  15.889 < 2e-16
APR.MDC.Code20 -7.396e-02  6.288e-02  -1.176 0.239499
APR.MDC.Code21 -2.739e-01  3.489e-02  -7.849 4.28e-15
APR.MDC.Code22 1.133e-01  3.024e-01  0.375 0.707983
APR.MDC.Code23 8.237e-02  2.585e-02  3.187 0.001439
APR.MDC.Code24 3.989e-02  4.602e-02  0.867 0.386102
APR.MDC.Code25 -4.578e-01  2.142e-01  -2.138 0.032551
APR.Severity.of.Illness.Code2 2.189e-01  7.852e-03  27.877 < 2e-16
APR.Severity.of.Illness.Code3 5.630e-01  1.120e-02  50.281 < 2e-16
APR.Severity.of.Illness.Code4 1.186e+00  1.922e-02  61.714 < 2e-16
APR.Risk.of.MortalityMajor -1.735e-01  1.949e-02  -8.904 < 2e-16
APR.Risk.of.MortalityMinor -3.209e-01  2.269e-02 -14.144 < 2e-16
APR.Risk.of.MortalityModerate -2.459e-01  2.124e-02 -11.573 < 2e-16
APR.Medical.Surgical.DescriptionSurgical 6.980e-01  8.401e-03  83.093 < 2e-16
Payment.Typeology.IFederal/State/Local/VA 2.174e-01  1.168e-01  1.862 0.062645
Payment.Typeology.IManaged Care, Unspecified 8.569e-02  3.055e-02  2.805 0.005040
Payment.Typeology.IMedicaid 3.521e-02  1.302e-02  2.704 0.006849
Payment.Typeology.IMedicare -2.395e-02  1.494e-02  -1.603 0.108977

```

Payment.Typeology.2Miscellaneous/other	5.624e-01	1.420e-01	3.960	7.52e-05
Payment.Typeology.2Private Health Insurance	1.976e-03	1.678e-02	0.118	0.906260
Payment.Typeology.2Self-Pay	6.928e-01	1.230e-01	5.633	1.78e-08
Payment.Typeology.3Federal/State/Local/VA	1.948e-02	2.455e-01	0.079	0.936746
Payment.Typeology.3Managed Care, Unspecified	1.084e-01	2.848e-01	0.381	0.703568
Payment.Typeology.3Medicaid	3.454e-02	9.011e-02	0.383	0.701490
Payment.Typeology.3Medicare	1.077e-01	1.286e-01	0.838	0.401944
Payment.Typeology.3Miscellaneous/other	-6.365e-01	1.514e-01	-4.203	2.64e-05
Payment.Typeology.3Private Health Insurance	2.361e-01	1.068e-01	2.212	0.027003
Payment.Typeology.3Self-Pay	7.310e-02	8.971e-02	0.815	0.415136
Birth.Weight	-2.144e-04	1.057e-05	-20.281	< 2e-16
Emergency.Department.IndicatorY	-1.607e-03	4.372e-02	-0.037	0.970679

```

(Intercept)          ***
Age.Group18 to 29    ***
Age.Group30 to 49    ***
Age.Group50 to 69    ***
Age.Group70 or Older ***
GenderM              ***
RaceOther Race      **
RaceWhite            **
EthnicitySpanish/Hispanic
Type.of.AdmissionEmergency ***
Type.of.AdmissionNewborn ***
Type.of.AdmissionUrgent ***
Patient.DispositionCourt/Law Enforcement
Patient.DispositionExpired
Patient.DispositionFacility w/ Custodial/Supportive Care
Patient.DispositionFederal Health Care Facility **
Patient.DispositionHome or Self Care **
Patient.DispositionHome w/ Home Health Services
Patient.DispositionHospice - Home .
Patient.DispositionHospice - Medical Facility
Patient.DispositionInpatient Rehabilitation Facility
Patient.DispositionLeft Against Medical Advice ***
Patient.DispositionMedicare cert Long Term Care Hospital
Patient.DispositionPsychiatric Hospital or Unit of Hosp *
Patient.DispositionShort-term Hospital .
Patient.DispositionSkilled Nursing Home
CCS.Diagnosis.Code
CCS.Procedure.Code ***
APR.MDC.Code2        *
APR.MDC.Code3        ***
APR.MDC.Code4        ***
APR.MDC.Code5        ***
APR.MDC.Code6        ***
APR.MDC.Code7        ***
APR.MDC.Code8        **
APR.MDC.Code9        ***
APR.MDC.Code10       ***
APR.MDC.Code11       ***
APR.MDC.Code12       ***
APR.MDC.Code13       ***
APR.MDC.Code14       ***
APR.MDC.Code15       ***
APR.MDC.Code16       ***
APR.MDC.Code17       ***
APR.MDC.Code18       ***
APR.MDC.Code19       ***
APR.MDC.Code20       ***
APR.MDC.Code21       ***
APR.MDC.Code22       ***
APR.MDC.Code23       **
APR.MDC.Code24       *
APR.MDC.Code25       *
APR.Severity.of.Illness.Code2 ***
APR.Severity.of.Illness.Code3 ***
APR.Severity.of.Illness.Code4 ***
APR.Risk.of.MortalityMajor ***
APR.Risk.of.MortalityMinor ***
APR.Risk.of.MortalityModerate ***
APR.Medical.Surgical.DescriptionSurgical
Payment.Typeology.1Federal/State/Local/VA .
Payment.Typeology.1Managed care, unspecified **
Payment.Typeology.1Medicaid **
Payment.Typeology.1Medicare
Payment.Typeology.1Miscellaneous/other
Payment.Typeology.1Private Health Insurance
Payment.Typeology.1Self-Pay
Payment.Typeology.2Federal/State/Local/VA ***
Payment.Typeology.2Managed Care, Unspecified **
Payment.Typeology.2Medicaid
Payment.Typeology.2Medicare
Payment.Typeology.2Miscellaneous/other ***
Payment.Typeology.2Private Health Insurance
Payment.Typeology.2Self-Pay ***
Payment.Typeology.3Federal/State/Local/VA
Payment.Typeology.3Managed Care, Unspecified
Payment.Typeology.3Medicaid
Payment.Typeology.3Medicare
Payment.Typeology.3Miscellaneous/other ***
Payment.Typeology.3Private Health Insurance
Payment.Typeology.3Self-Pay ***
Birth.Weight
Emergency.Department.IndicatorY
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 0.604 on 41831 degrees of freedom  
Multiple R-squared: 0.6273, Adjusted R-squared: 0.6266  
F-statistic: 869.3 on 81 and 41831 DF, p-value: < 2.2e-16

```

> pred.lin <- predict(model.lin, newdata=test)
> SSE.lin <- sum((pred.lin - test$log.Total.Costs)^2)
> train.mean <- mean(train$log.Total.Costs)
> SST.lin <- sum((train.mean - test$log.Total.Costs)^2)
> OSR2.lin <- 1 - SSE.lin/SST.lin
> OSR2.lin
[1] 0.6288558

```



### R Output 3: Linear Regression for log.Length.of.Stay

Call:

```
lm(formula = logLength.of.Stay ~ . - Total.Costs - Factors -  
    Length.of.Stay, data = train)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.9471	-0.3771	-0.0065	0.3606	3.8410

Coefficients:

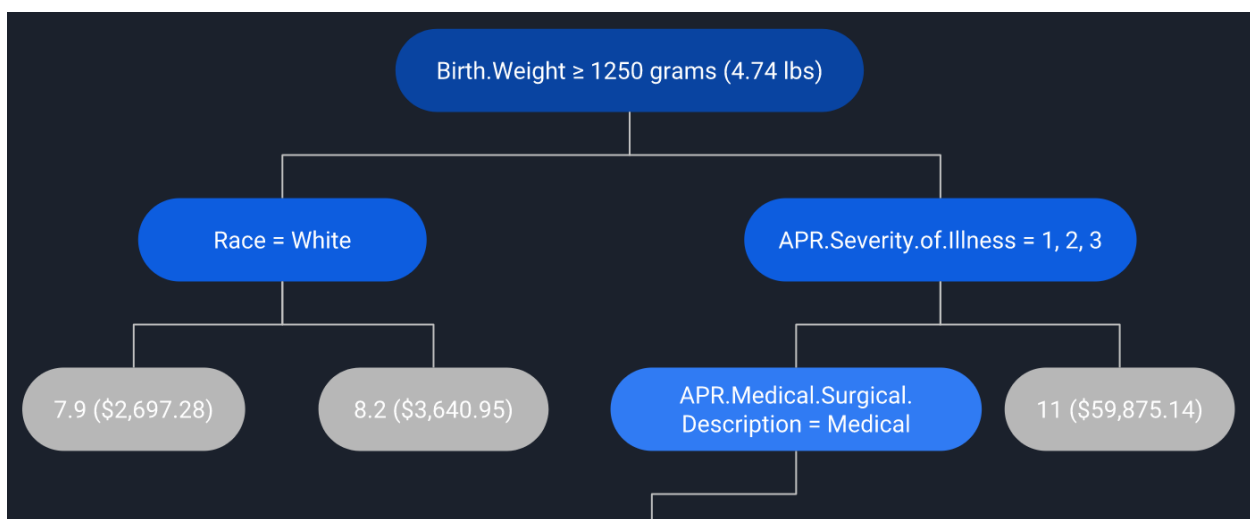
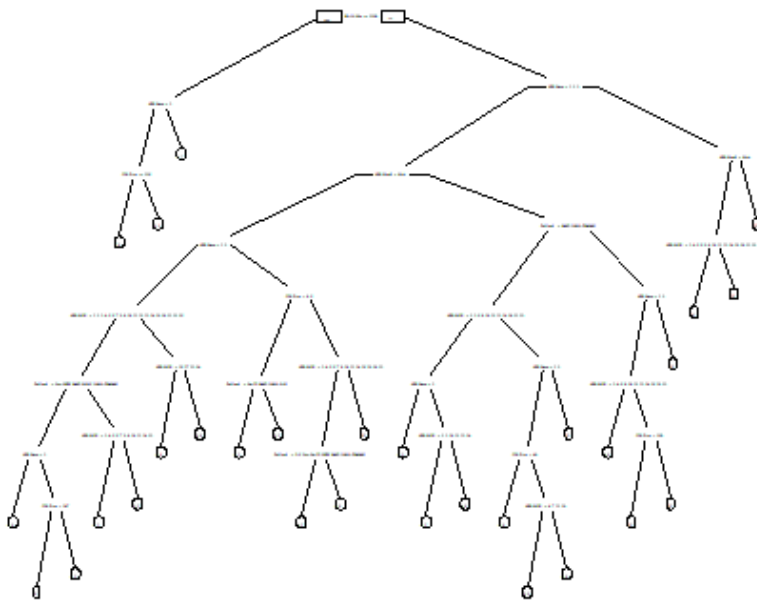
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.288e+00	2.627e-01	8.710	< 2e-16 ***
Age.Group18 to 29	1.181e-01	1.759e-02	6.712	1.95e-11 ***
Age.Group30 to 49	9.478e-02	1.600e-02	5.924	3.17e-09 ***
Age.Group50 to 69	2.394e-02	1.564e-02	1.531	0.125813
Age.Group70 or Older	-1.269e-01	1.871e-02	-6.781	1.21e-11 ***
GenderM	3.978e-03	6.856e-03	0.580	0.561731
RaceOther Race	-3.969e-02	8.973e-03	-4.424	9.73e-06 ***
RaceWhite	-3.576e-02	9.488e-03	-3.769	0.000164 ***
EthnicitySpanish/Hispanic	-1.138e-02	4.449e-01	-0.026	0.979602
Type.of.AdmissionEmergency	2.242e-01	4.671e-02	4.801	1.58e-06 ***
Type.of.AdmissionNewborn	5.017e-01	5.651e-02	8.878	< 2e-16 ***
Type.of.AdmissionUrgent	2.926e-01	1.166e-02	25.097	< 2e-16 ***
Patient.DispositionCourt/Law Enforcement	2.585e-02	3.147e-01	0.082	0.934530
Patient.DispositionExpired	-5.186e-01	1.834e-01	-2.827	0.004698 **
Patient.DispositionFacility w/ Custodial/Supportive Care	-2.297e-01	1.970e-01	-1.166	0.243753
Patient.DispositionFederal Health Care Facility	-9.500e-01	4.060e-01	-2.340	0.019274 *
Patient.DispositionHome or Self Care	-6.575e-01	1.821e-01	-3.611	0.000305 ***
Patient.DispositionHome w/ Home Health Services	-1.344e-01	1.821e-01	-0.738	0.460448
Patient.DispositionHospice - Home	-2.750e-01	1.876e-01	-1.466	0.142699
Patient.DispositionHospice - Medical Facility	-2.353e-01	1.863e-01	-1.263	0.206539
Patient.DispositionInpatient Rehabilitation Facility	1.269e-01	1.840e-01	0.690	0.490405
Patient.DispositionLeft Against Medical Advice	-9.066e-01	1.861e-01	-4.871	1.12e-06 ***
Patient.DispositionMedicare Cert Long Term Care Hospital	-4.851e-03	1.928e-01	-0.025	0.979932
Patient.DispositionPsychiatric Hospital or Unit of Hosp	-3.386e-01	1.932e-01	-1.753	0.079694 .
Patient.DispositionShort-term Hospital	-4.316e-01	1.869e-01	-2.309	0.020936 *
Patient.DispositionSkilled Nursing Home	1.823e-01	1.823e-01	1.000	0.317493
CCS.Diagnosis.Code	5.351e-05	5.145e-05	1.040	0.298390
CCS.Procedure.Code	3.963e-04	4.460e-05	8.886	< 2e-16 ***
APR.MDC.Code2	-1.285e-01	9.279e-02	-1.385	0.166153
APR.MDC.Code3	-1.237e-01	2.936e-02	-4.214	2.51e-05 ***
APR.MDC.Code4	-1.771e-02	1.963e-02	-0.902	0.366927
APR.MDC.Code5	-2.370e-01	1.603e-02	-14.786	< 2e-16 ***
APR.MDC.Code6	1.503e-01	1.743e-02	8.625	< 2e-16 ***
APR.MDC.Code7	8.445e-02	2.237e-02	3.775	0.000160 ***
APR.MDC.Code8	-1.270e-01	1.895e-02	-6.703	2.06e-11 ***
APR.MDC.Code9	-4.356e-02	2.732e-02	-1.595	0.110795
APR.MDC.Code10	-1.216e-01	2.363e-02	-5.147	2.66e-07 ***
APR.MDC.Code11	1.452e-02	2.135e-02	0.680	0.496475
APR.MDC.Code12	-4.725e-01	3.543e-02	-13.338	< 2e-16 ***
APR.MDC.Code13	-9.490e-02	2.977e-02	-3.188	0.001436 **
APR.MDC.Code14	2.379e-01	1.936e-02	12.292	< 2e-16 ***
APR.MDC.Code15	3.082e-01	6.665e-02	4.624	3.78e-06 ***
APR.MDC.Code16	2.349e-01	2.933e-02	8.008	1.20e-15 ***
APR.MDC.Code17	6.248e-01	2.623e-02	23.822	< 2e-16 ***
APR.MDC.Code18	2.356e-02	2.132e-02	1.105	0.269018
APR.MDC.Code19	1.258e+00	3.570e-02	35.249	< 2e-16 ***
APR.MDC.Code20	3.001e-01	6.396e-02	4.692	2.71e-06 ***
APR.MDC.Code21	-1.823e-01	3.651e-02	-4.993	5.98e-07 ***
APR.MDC.Code22	1.215e-01	2.816e-01	0.432	0.666104
APR.MDC.Code23	7.201e-01	2.689e-02	26.782	< 2e-16 ***
APR.MDC.Code24	1.475e-01	4.752e-02	3.105	0.001904 **
APR.MDC.Code25	-1.530e-01	1.995e-01	-0.767	0.442938
Payment.Typology.2Self-Pay	-1.445e-01	7.764e-02	-1.861	0.062692 .
Payment.Typology.3Federal/State/Local/VA	-4.112e-01	2.299e-01	-1.789	0.073666 .
Payment.Typology.3Managed Care, Unspecified	-3.448e-01	2.966e-01	-1.162	0.245077
Payment.Typology.3Medicaid	-1.735e-01	9.473e-02	-1.832	0.066997 .
Payment.Typology.3Medicare	-1.373e-01	1.345e-01	-1.021	0.307152
Payment.Typology.3None	-7.647e-01	1.693e-01	-4.517	6.29e-06 ***
Payment.Typology.3Private Health Insurance	1.927e-01	1.116e-01	1.727	0.084148 .
Payment.Typology.3Self-Pay	-1.183e-01	9.427e-02	-1.255	0.209374
Birth.Weight	-1.043e-04	1.090e-05	-9.572	< 2e-16 ***
Emergency.Department.IndicatorY	5.040e-02	4.632e-02	1.088	0.276570

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

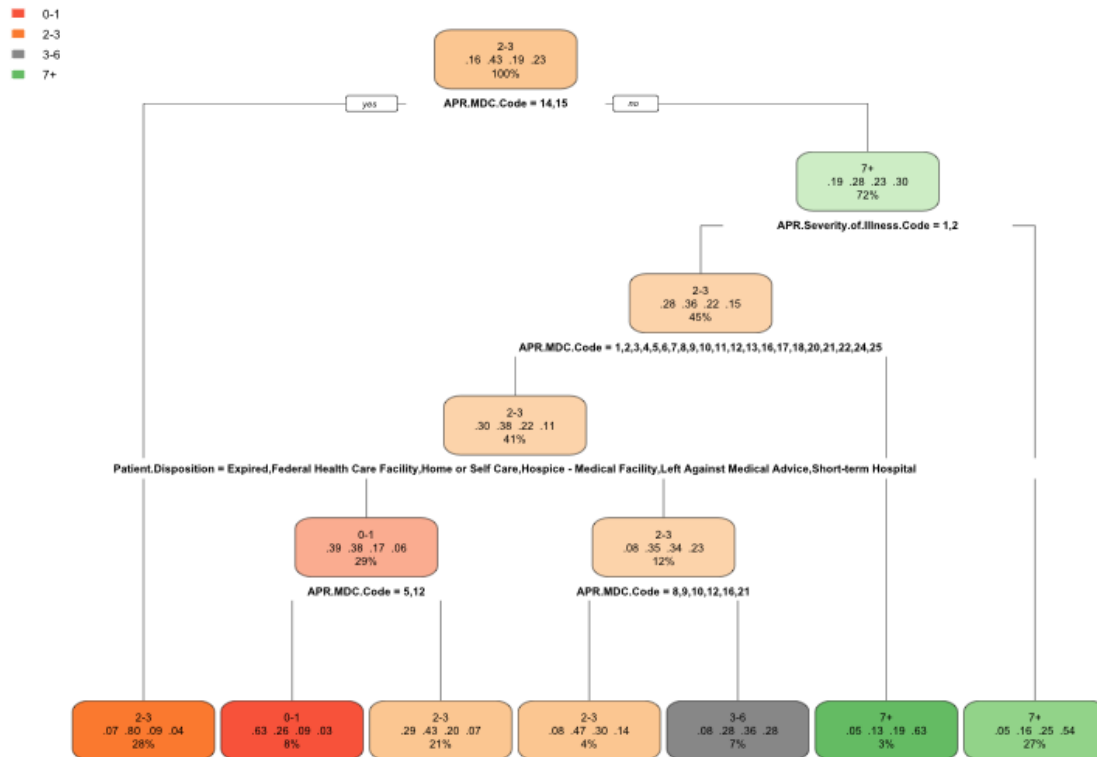
Residual standard error: 0.6286 on 41830 degrees of freedom  
Multiple R-squared: 0.5129, Adjusted R-squared: 0.512  
F-statistic: 537.2 on 82 and 41830 DF, p-value: < 2.2e-16

#### R Output 4: Regression Tree for log.Total.Costs

```
> library(rpart)
> library(rpart.plot)
> tree500 <- rpart(log.Total.Costs ~ ., data=train, method="anova", minbucket=500, cp=0)
> plotcp(tree500)
> tree501 <- rpart(log.Total.Costs ~ ., data = train, method="anova", minbucket=500, cp=0.00080008)
> prp(tree501)
> tree.model <- tree501
> pred.tree <- predict(tree.model, newdata=test)
> SSE.tree <- sum((pred.tree - test$log.Total.Costs)^2)
> SST.tree <- sum((train.mean - test$log.Total.Costs)^2)
> OSR2.tree <- 1 - SSE.tree/SST.tree
> OSR2.tree - OSR2.lin
[1] 0.00223996
> OSR2.tree
[1] 0.6310958
```



## R Output 5: Classification Tree for Length.of.Stay.Categories



> Confusion

pred

	0-1	2-3	4-6	7+
0-1	699	1246	68	202
2-3	339	4631	285	700
4-6	111	1151	344	1030
7+	34	443	279	2410

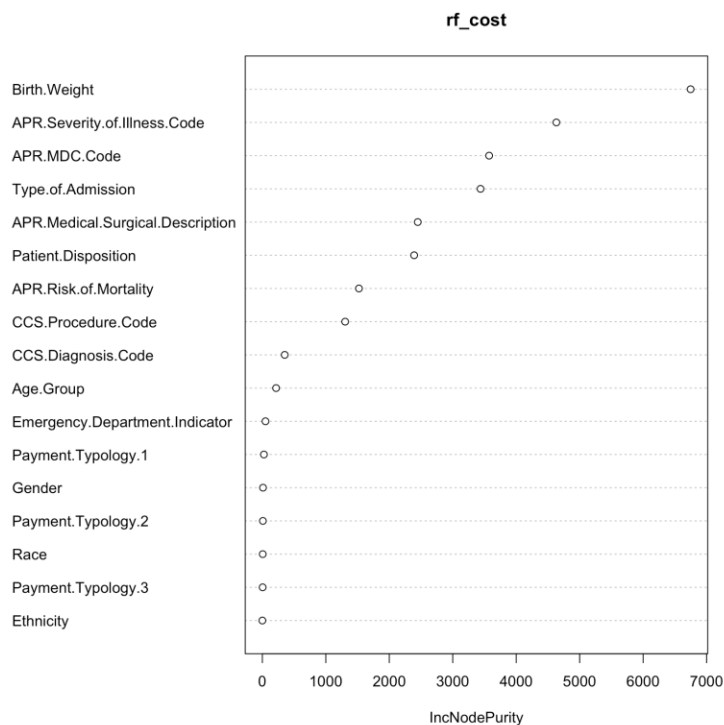
> (Confusion[1,1]+Confusion[2,2]+Confusion[3,3]+Confusion[4,4])/sum(Confusion)

[1] 0.5785857

## R Output 6: Random Forest for log.Total.Costs

Only 1 TuneRF was used due to the time consuming nature of the function and the similarity of the trees.

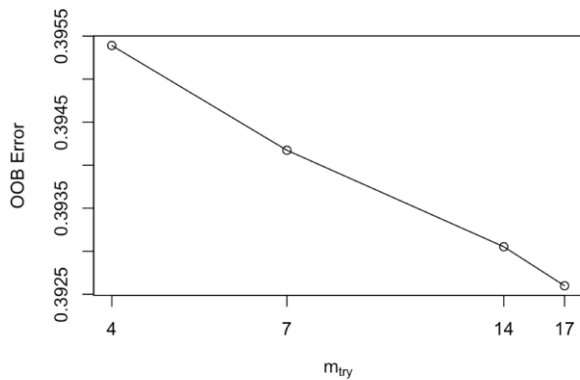
```
> start_time <- Sys.time()
> rf_cost = randomForest(Total.Costs~.- Length.of.Stay - Factors-logLength.of.Stay,
+                         data=train,ntree=500,nodesize=1000,mtry=7)
> end_time <- Sys.time()
> end_time-start_time
Time difference of 3.860957 mins
> varImpPlot(rf_cost)
```



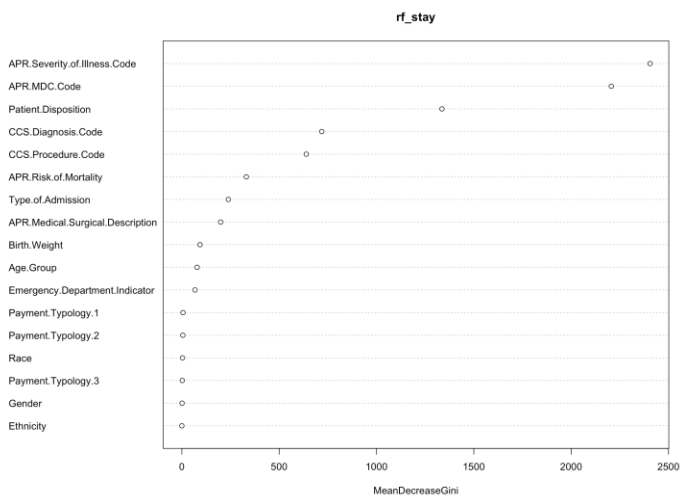
```
> pred_cost = predict(rf_ini,newdata=test)
> RFCost <- table(test$Factors,pred_cost)
> SSE_rf=sum((test$Total.Costs - pred_cost)^2)
> SST_rf = sum((test$Total.Costs - mean(train$Total.Costs))^2)
> OSR = 1 - SSE_rf/SST_rf
> OSR
[1] 0.6892182
```

## R Output 7: Random Forest for Length.of.Stay.Category

```
> y.dat <- train$Factors
> x.dat <- train
> x.dat$Length.of.Stay <- NULL
> x.dat$Total.Costs <- NULL
> x.dat$Factors <- NULL
> x.dat$logLength.of.Stay <- NULL
> set.seed(1760)
> tuneRF(x.dat, y.dat, mtryStart = 7, stepFactor = 2, ntreeTry=500,
+       nodesize=1000, improve=0.0001)
```



```
> rf_stay = randomForest(Factors~.- Length.of.Stay - Total.Costs-logLength.of.Stay,
+ data=train,ntree=500,nodesize=1000,mtry=10)
```



```
> RFConf
```

```
pred_rf
      0-1  2-3  4-6  7+
0-1    784 1245   72 114
2-3    294 5034  160 467
4-6     74 1445  237 880
7+      27  521  215 2403
```

```
> (RFConf[1,1]+RFConf[2,2]+RFConf[3,3]+RFConf[4,4])/sum(Confusion)
[1] 0.6053536
```

## R Output 8: Linear Discriminant Analysis for Length.of.Stay.Category

```
Call:
lda(Factors ~ . - Length.of.Stay - Total.Costs - logLength.of.Stay,
    data = train)

Prior probabilities of groups:
  0-1      2-3      4-6      7+
0.1585666 0.4262162 0.1886527 0.2265646

Group means:
Age.Group18 to 29 Age.Group30 to 49 Age.Group50 to 69 Age.Group70 or Older GenderM RaceOther Race
0-1 0.06364731 0.1739392 0.3549503 0.2451098 0.5470960 0.3901595
2-3 0.12544783 0.2576131 0.1783475 0.1329489 0.3560793 0.3622369
4-6 0.09194385 0.2053876 0.3319843 0.2741874 0.4434046 0.3851018
7+ 0.06434288 0.1573294 0.3927970 0.3029697 0.5113732 0.4017481

RaceWhite EthnicitySpanish/Hispanic Type.of.AdmissionEmergency Type.of.AdmissionNewborn
0-1 0.4637376 0.0000000000 0.2279567 0.06048751
2-3 0.4918831 0.0000559785 0.2369570 0.25626959
4-6 0.4122929 0.0000000000 0.4573163 0.03692930
7+ 0.3574136 0.0001053075 0.5064238 0.03327717

Type.of.AdmissionUrgent Patient.DispositionCourt/Law Enforcement Patient.DispositionExpired
0-1 0.12985254 0.0000000000 0.010382185
2-3 0.06140842 0.0001119570 0.004030452
4-6 0.13152902 0.0001264702 0.013279373
7+ 0.23946925 0.0003159225 0.040648694

Patient.DispositionFacility w/ Custodial/Supportive Care Patient.DispositionFederal Health Care Facility
0-1 0.0007523322 0.0001504664
2-3 0.0008396776 0.0000559785
4-6 0.0027823448 0.0000000000
7+ 0.0028433024 0.0001053075

Patient.DispositionHome or Self Care Patient.DispositionHome w/ Home Health Services
0-1 0.8976828 0.06184171
2-3 0.8390618 0.12012987
4-6 0.5577337 0.30491969
7+ 0.3176074 0.36404802

Patient.DispositionHospice - Home Patient.DispositionHospice - Medical Facility
0-1 0.001053265 0.001504664
2-3 0.001511420 0.001119570
4-6 0.004085868 0.003288226
7+ 0.012005055 0.018744735

Patient.DispositionInpatient Rehabilitation Facility Patient.DispositionLeft Against Medical Advice
0-1 0.0004513993 0.014444779
2-3 0.0044223018 0.005206001
4-6 0.0163146579 0.006070570
7+ 0.0349620893 0.003369840

Patient.DispositionMedicare Cert Long Term Care Hospital
0-1 0.0004513993
2-3 0.0006157635
4-6 0.0017705830
7+ 0.0071609099

Patient.DispositionPsychiatric Hospital or Unit of Hosp Patient.DispositionShort-term Hospital
0-1 0.001504664 0.005115859
2-3 0.001623377 0.002519033
4-6 0.002023523 0.003414696
7+ 0.004317607 0.011794440

Patient.DispositionSkilled Nursing Home CCS.Diagnosis.Code CCS.Procedure.Code APR.MDC.Code2 APR.MDC.Code3
0-1 0.00466446 136.5391 88.70990 0.0019560638 0.02512790
2-3 0.01869682 172.7913 126.98102 0.0008956561 0.01214734
4-6 0.08321740 152.6384 95.94246 0.0010117617 0.01100291
7+ 0.18112890 181.9901 103.08920 0.0010530750 0.01295282

APR.MDC.Code4 APR.MDC.Code5 APR.MDC.Code6 APR.MDC.Code7 APR.MDC.Code8 APR.MDC.Code9 APR.MDC.Code10
0-1 0.04468853 0.35073729 0.07598556 0.02347277 0.07102016 0.01745411 0.03746615
2-3 0.04103224 0.09404389 0.05933721 0.01953650 0.06342365 0.01556202 0.03101209
4-6 0.07094979 0.17769065 0.13051726 0.04274693 0.09131150 0.02453522 0.02149994
7+ 0.05370682 0.18481466 0.10309604 0.04770430 0.05539174 0.01326874 0.01663858

APR.MDC.Code11 APR.MDC.Code12 APR.MDC.Code13 APR.MDC.Code14 APR.MDC.Code15 APR.MDC.Code16 APR.MDC.Code17
0-1 0.01910924 0.044086669 0.029641890 0.05808005 0.06861270 0.01128498 0.01865784
2-3 0.02871697 0.003638603 0.013938648 0.26343484 0.25845275 0.01057994 0.00559785
4-6 0.06121158 0.002023523 0.012899962 0.09308208 0.03781459 0.02263817 0.02668522
7+ 0.04981045 0.002211457 0.006529065 0.01316344 0.03643639 0.01821820 0.04538753

APR.MDC.Code18 APR.MDC.Code19 APR.MDC.Code20 APR.MDC.Code21 APR.MDC.Code22 APR.MDC.Code23 APR.MDC.Code24
0-1 0.01218778 0.002106530 0.001805597 0.012789648 0.0000000000 0.009027987 0.0009027987
2-3 0.01897671 0.005485893 0.002519033 0.006829378 0.0001119570 0.006381549 0.0026869682
4-6 0.05159985 0.030985203 0.003920577 0.009105856 0.0001264702 0.008599975 0.0072088023
7+ 0.08887953 0.090143218 0.004106992 0.010109520 0.0002106150 0.064764111 0.0090564448

APR.MDC.Code25 APR.MDC.Code26 APR.MDC.Code27 APR.MDC.Code28 APR.MDC.Code29 APR.MDC.Code30 APR.MDC.Code31
0-1 0.0001504664 0.3972314 0.07824255 0.01173638
2-3 0.0000559785 0.3560233 0.11307658 0.01242723
4-6 0.0002529404 0.4191223 0.32970785 0.05792336
7+ 0.0006318450 0.2555813 0.42786436 0.25526537

APR.Risk.of.MortalityMajor APR.Risk.of.MortalityMinor APR.Risk.of.MortalityModerate
0-1 0.04092687 0.7172735 0.2309660
2-3 0.05446708 0.7936073 0.1443126
4-6 0.18793474 0.4690780 0.3078285
7+ 0.30328559 0.2601095 0.2681129

APR.Medical.Surgical.DescriptionSurgical Payment.Typeology.1Federal/State/Local/VA
0-1 0.5740295 0.0009027987
2-3 0.2834751 0.0004478280
4-6 0.4420134 0.0007588213
7+ 0.4036436 0.0012636900

Payment.Typeology.1Managed Care, Unspecified Payment.Typeology.1Medicaid Payment.Typeology.1Medicare
0-1 0.007974722 0.2347277 0.3551008
2-3 0.010859830 0.2822436 0.1963166
4-6 0.012267611 0.2413052 0.4091311
7+ 0.014848357 0.2687447 0.4837826
```



Payment.Typeology.1Miscellaneous/Other	Payment.Typeology.1Private Health Insurance	Payment.Typeology.1Self-Pay
0-1 0.010081252	0.2467650	0.012338249
2-3 0.004814151	0.3640842	0.013602777
4-6 0.005058809	0.2158847	0.011002909
7+ 0.007160910	0.1367944	0.007476832
Payment.Typeology.2Blue Cross/Blue Shield	Payment.Typeology.2Federal/State/Local/VA	
0-1 0.06304544	0.0006018658	
2-3 0.03330721	0.0002239140	
4-6 0.06159100	0.0010117617	
7+ 0.07213564	0.0015796125	
Payment.Typeology.2Managed Care, Unspecified	Payment.Typeology.2Medicaid	Payment.Typeology.2Medicare
0-1 0.001504664	0.2715919	0.14459825
2-3 0.003470667	0.3228280	0.07243618
4-6 0.002276464	0.3097256	0.14278487
7+ 0.001474305	0.3277169	0.18133951
Payment.Typeology.2Miscellaneous/Other	Payment.Typeology.2Private Health Insurance	Payment.Typeology.2Self-Pay
0-1 0.0000000000	0.09223593	0.4261210
2-3 0.0000000000	0.05301164	0.5117555
4-6 0.0000000000	0.09839383	0.3821930
7+ 0.0001053075	0.10467565	0.3099200
Payment.Typeology.3Federal/State/Local/VA	Payment.Typeology.3Managed Care, Unspecified	
0-1 0.0006018658	0.0003009329	
2-3 0.0002239140	0.0000559785	
4-6 0.0001264702	0.0001264702	
7+ 0.0000000000	0.0001053075	
Payment.Typeology.3Medicaid	Payment.Typeology.3Medicare	Payment.Typeology.3None
0-1 0.06319591	0.0012037316	0.4264219
2-3 0.04265562	0.0005038065	0.5147223
4-6 0.07879094	0.0010117617	0.3839636
7+ 0.09898905	0.0018955350	0.3091828
Payment.Typeology.3Private Health Insurance	Payment.Typeology.3Self-Pay Birth.Weight	
0-1 0.0021065302	0.5058682	231.7936
2-3 0.0007277206	0.4404389	848.2423
4-6 0.0029088150	0.5318073	113.6461
7+ 0.0081086773	0.5795072	103.0960
Emergency.Department.Indicator		
0-1 0.2279567		
2-3 0.2362293		
4-6 0.4560516		
7+ 0.5046335		

Coefficients of linear discriminants:

	LD1	LD2	LD3
Age.Group18 to 29	2.217165e-01	0.3737828840	-0.2531747471
Age.Group30 to 49	1.587879e-01	-0.3398962591	-0.1111745701
Age.Group50 to 69	3.393714e-02	-0.2401701800	0.0105190552
Age.Group70 or Older	-2.781109e-01	-0.1443487495	-0.2449102480
GenderM	3.496184e-02	0.1497062589	0.1382555064
RaceOther Race	-6.456375e-02	0.0304754000	-0.0298268227
RaceWhite	-3.123183e-02	0.0196560517	-0.0352871829
EthnicitySpanish/Hispanic	6.727426e-01	-1.1296453693	3.0934305364
Type.of.AdmissionEmergency	3.583579e-01	-0.5762331340	-0.3361647292
Type.of.AdmissionNewborn	3.916505e-01	-3.3832378907	-0.7787807800
Type.of.AdmissionUrgent	5.571684e-01	-0.4178465166	0.0257068195
Patient.DispositionCourt/Law Enforcement	1.021742e-01	-0.0309966000	0.4315403203
Patient.DispositionExpired	-6.720499e-01	0.5787968268	-0.2850269351
Patient.DispositionFacility w/ Custodial/Supportive Care	-1.507296e-01	0.1424971721	-0.7615573749
Patient.DispositionFederal Health Care Facility	-1.591404e+00	0.9257941317	2.0809760547
Patient.DispositionHome or Self Care	-1.021266e+00	0.5641563472	0.1531018541
Patient.DispositionHome w/ Home Health Services	7.565961e-02	-0.2530552256	-0.3695949528
Patient.DispositionHospice - Home	-5.162335e-02	0.3233110221	0.5091058717
Patient.DispositionHospice - Medical Facility	-4.814427e-02	0.3823250134	0.9532680952
Patient.DispositionInpatient Rehabilitation Facility	6.832951e-01	-0.1670807227	0.4716729682
Patient.DispositionLeft Against Medical Advice	-1.495609e+00	1.3277383652	0.4901104829
Patient.DispositionMedicare Cert Long Term Care Hospital	1.804115e-01	0.2211174639	1.1226905252
Patient.DispositionPsychiatric Hospital or Unit of Hosp	-5.639498e-01	0.1165259756	0.5140297782
Patient.DispositionShort-term Hospital	-7.674899e-01	0.4663254478	0.3857721182
Patient.DispositionSkilled Nursing Home	6.908522e-01	-0.0720509762	0.3441872617
CCS.Diagnosis.Code	-7.408302e-05	0.0005850426	0.0018855891
CCS.Procedure.Code	9.396929e-04	-0.0004146049	0.0002799532
APR.MDC.Code2	-2.413206e-01	0.1901654953	0.5881266121
APR.MDC.Code3	-2.621355e-01	-0.0188502635	0.7323151871
APR.MDC.Code4	-1.028548e-01	-0.2563087814	-0.5407717301
APR.MDC.Code5	-3.506209e-01	0.6225502857	0.3360025923
APR.MDC.Code6	3.080375e-01	-0.3758615313	-0.6810192323
APR.MDC.Code7	1.608189e-01	-0.3238041533	-0.5341012924
APR.MDC.Code8	-3.966072e-01	-0.6440043594	-0.6993313841
APR.MDC.Code9	-1.661857e-01	-0.3278057154	-0.6131254776
APR.MDC.Code10	-3.826340e-01	-0.5936960821	-0.5818644661
APR.MDC.Code11	-1.388908e-02	-0.4980529546	-0.8282070699
APR.MDC.Code12	-6.391206e-01	2.1207803808	1.5807366344
APR.MDC.Code13	-2.168681e-01	-0.1986681621	0.3007115539
APR.MDC.Code14	-1.300555e-01	-2.4424384391	-0.0152842135

APR.MDC.Code15	9.844637e-01	1.0333415664	0.6367791645
APR.MDC.Code16	4.424958e-01	-0.2836092545	-0.6565107529
APR.MDC.Code17	1.098184e+00	0.0257684683	0.1750852748
APR.MDC.Code18	1.458287e-01	-0.1192520750	-0.2329864448
APR.MDC.Code19	2.827530e+00	-0.7820663957	0.1404598657
APR.MDC.Code20	7.121420e-01	-0.8777618378	-1.1006419059
APR.MDC.Code21	-3.433130e-01	0.0981526462	-0.0354238918
APR.MDC.Code22	2.318713e-01	-0.9382955213	-0.1760643662
APR.MDC.Code23	1.423204e+00	-0.1750678250	2.0872836642
APR.MDC.Code24	2.009795e-01	-0.2228732698	-0.6022329885
APR.MDC.Code25	3.762118e-01	0.1388858250	0.6990038002
APR.Severity.of.Illness.Code2	4.483410e-01	-0.2979421737	-0.6962993835
APR.Severity.of.Illness.Code3	1.372675e+00	-0.3316455442	-0.6448479933
APR.Severity.of.Illness.Code4	1.914554e+00	-0.2067166558	0.8862808577
APR.Risk.of.MortalityMajor	-2.277784e-01	-0.0673802049	-0.9460091563
APR.Risk.of.MortalityMinor	-8.402343e-01	0.2930435033	-1.0368806513
APR.Risk.of.MortalityModerate	-5.732492e-01	0.1209164753	-1.2987453473
APR.Medical.Surgical.DescriptionSurgical	5.547038e-01	0.2072000211	-0.2133339115
Payment.Typeology.1Federal/State/Local/VA	1.011948e-01	0.3589793846	0.3941078849
Payment.Typeology.1Managed Care, Unspecified	2.058080e-01	-0.2260071046	0.1114854463
Payment.Typeology.2Medicaid	6.899848e-02	0.0299088536	0.2019344482
Payment.Typeology.2Medicare	-8.566941e-02	0.1582382733	0.0738528257
Payment.Typeology.2Miscellaneous/Other	-1.373997e-01	0.3291399393	0.6556688959
Payment.Typeology.1Private Health Insurance	-1.187114e-02	-0.0205708322	-0.0581884225
Payment.Typeology.1Self-Pay	-3.897209e-02	-0.0746356523	-0.0498263111
Payment.Typeology.2Blue Cross/Blue Shield	-3.128045e-01	1.9364803400	-1.6980146756
Payment.Typeology.2Federal/State/Local/VA	1.594097e-01	1.9561333232	-1.7353103570
Payment.Typeology.2Managed Care, Unspecified	-2.341010e-01	1.7144772246	-1.9090630756
Payment.Typeology.2Medicaid	-4.469177e-01	2.0213412999	-1.8149654371
Payment.Typeology.2Medicare	-2.818922e-01	1.9999480811	-1.5632710521
Payment.Typeology.2Miscellaneous/Other	-6.886900e-01	2.0761594689	-0.0987230161
Payment.Typeology.2Private Health Insurance	-3.266830e-01	1.9228568708	-1.8505316391
Payment.Typeology.2Self-Pay	1.368494e-01	1.8773872752	-0.4359615643
Payment.Typeology.3Federal/State/Local/VA	-1.084889e+00	0.7396230536	0.5823130455
Payment.Typeology.3Managed Care, Unspecified	-3.686649e-01	1.3322144264	0.3443198401
Payment.Typeology.3Medicaid	-4.645697e-01	0.5313975095	-0.2248250607
Payment.Typeology.3Medicare	-1.983493e-01	0.7988234163	0.1396606454
Payment.Typeology.3None	-8.460577e-01	0.5971717332	-1.6097472260
Payment.Typeology.3Private Health Insurance	3.050822e-01	0.5978398085	-0.1004210876
Payment.Typeology.3Self-Pay	-3.627398e-01	0.4166650293	-0.2086833580
Birth.Weight	-3.273476e-04	-0.0001872310	0.0001216518
Emergency.Department.IndicatorY	1.172665e-01	-0.0488049992	-0.1300021408

Proportion of trace:

LD1 LD2 LD3

0.7658 0.1907 0.0436