What is the True Cost to stay in the Hospital?

The Honors Program
Senior Capstone Project
Student's Name: Samantha Alicandro
Faculty Sponsor: Alan Olinsky
April 2016

Table of Contents

ABSTRACT

Currently, the unfortunate reality that receiving diverse health procedures can be extremely expensive is widely acknowledged and woefully accepted. However, have you inquired or been curious about the specific factors that influence the cost per day expensed by a hospital? Through examination, investigation, and evaluation operating SAS Enterprise Guide, SAS Enterprise Miner and Tableau I have attempted to arrive at a conclusion for this very question. Utilizing a 1.5 million row data set provided by Rhode Island, for the years 2003-2013, I analyzed the assorted elements conceivably bearing impact on the cost per day at a hospital. Regressions, decision trees, neural networks, ANOVA, linear models, and a countless number of visual representations genuinely assisted in attaining a robust conclusion. Overall, a number of various input variables, with unique magnitudes, such as age, services provided, year of discharge, and hospital provider sincerely shape the overarching cost per day at a hospital.

BACKGROUND RESEARCH

Introduction

Healthcare is an extremely “hot topic” in society today. I believe that Healthcare is truly a fascinating subject, and one that everyone should be interested in, for based on law, it is required that all citizens possess healthcare. It is extremely important, according to government officials and hospital representatives, to be aware of your healthcare, and the fees hospitals are currently charging for your own knowledge and awareness in case of an emergency. This is not only a social, but political and economic debate amongst all individuals that easily sparks controversy. In this capstone, I will be utilizing quantitative data to analyze the various factors impacting the cost per day of healthcare charges by hospitals in Rhode Island.

In the data set I possess, I will be analyzing inpatient data from fourteen distinct hospitals in Rhode Island for the calendar years 2003 and 2013. Variables such as length of stay, expected source of payment, reason for admission, disposition, and more are encompassed by this over 1.5 million record data set. My overarching goal will be to analyze the various and diverse variables in SAS Enterprise Guide and SAS Enterprise Miner to arrive at a conclusion illustrating the “average” cost per day of a hospital stay. In this study the cost per day is the cost charged, or fees incurred, by each distinct hospital per day. Statistical tools such as summary statistics, ANOVA, one-way frequencies, regressions, decision trees and more will be utilized to fully examine and explore the hospital data set. I will be comparing my findings to supporting outside research, demonstrating both descriptive and quantitative data to back up my research.

Background on Healthcare

"Healthcare Crisis History" an article by PBS describes the changes in healthcare over the past century, separating major events by decade. It is extremely important that we understand the past of healthcare before jumping into current debates and analysis of where healthcare stands today.

In 1900, the American Medical Association became a nationally recognized and robust force beginning a period of what is called "organized medicine." In addition, at this point in history surgeries were becoming much more common and doctors began charging patients for services and procedures, prior to this it was free. Moreover, at this point in time not everyone had insurance to pay for these services so companies, such as railroads, began what is now known as medical insurance, or medical programs. In 1910, hospitals became a sanitary and clean place for the sick and to perform surgeries. The controversial topic of health insurance was put on hold due to the war. In the 1920s, more companies, such as General Motors, began contracting with insurance companies to insure the workers. Furthermore, the "stereotype" of doctors arose in this time period, creating a fascination and prestige around those in the profession. At this time, doctors rapidly became the elite of society ("Healthcare Crisis History").

In the 1930s, there was still and intense push to create health insurance however, by law nothing had been instated yet. Against the advice of countless professionals Blue Cross began offering private coverage for hospital care in a variety of states. In the 1940s, Presidents Roosevelt and Truman attempted to take action. Roosevelt proposed an "economic bill of rights" which included healthcare, and Truman proposed a national health program plan. A variety of new medications to treat diseases and infections, along with vaccines, were available in the 1950s ("Healthcare Crisis History").

In the 1960s the cost of a hospital stay nearly doubled, causing many elderly people to have substantial difficulty affording insurance. At this point in time there were over 700 insurance companies, selling private insurance to individuals. In addition, there was a shortage of doctors, hence the need to expand the education in the health profession. Also, Medicare and Medicaid were now law. In the 1970s, healthcare costs skyrocketed again, for the government was not prepared to take on the costs of Medicare and Medicaid. Women began entering the medical profession as well, commonly as nurses. There were no significant changes in the 1980s, however healthcare shifted towards the private sector under insurance ("Healthcare Crisis History").

In the 1990s and 2000s there were significant changes to healthcare and the healthcare system. First, healthcare costs began rising at nearly double the rate of inflation! Because of the extreme costs, the healthcare system expanded significantly. At this point in time 16% of Americans did not possess health insurance, because of the high cost. In addition, the Human Genome Project, to discover all of humans DNA, commenced. The project was finally completed in 2003. Moreover, Medicare needed to be altered, for the present condition and costs were not sustainable by the government. In addition, pharmaceutical and medical device companies began advertising directly to consumers. Once again, healthcare was on the rise. Today, healthcare is required and continues to steady increase in cost, however it continues to be a controversial and heavily discussed economic, social, and political issue ("Healthcare Crisis History").

New York Times Articles

A common trend throughout the decades was the increasing cost of healthcare. My project focuses on analyzing the cost of healthcare because of its prominence throughout American history. From my annotative bibliography, I have chosen five main articles to focus on, for the context and information encompassed in the articles and scholarly sources. I do believe, however cannot guarantee, that these sources will correlate directly to the Hospital data analysis. Three of the articles, “When a Hospital Stay Is Not a Stay”, “Hospital Billing Varies Wildly, Government Data Shows”, and “When Hospitals Buy Doctors’ Offices, and Patient Fees Soar” are all published as articles in The New York Times. Although each article may argue diverse topics in the healthcare field, they highly correlate to the overarching concept of average cost to stay in a hospital.

“When a Hospital Stay Is Not a Stay” is a recent article discussing the impact of patient status, Medicare, and appealing charges of a hospital stay. Unfortunately, there is no federal law requiring hospitals to inform a patient of their status. Moreover, being an inpatient or an outpatient significantly effects the cost of the hospital stay for different rates are allocated based on patient status. The article pronounces, “It turned out that she had not been admitted as an impatient, he said, so Medicare would not cover the cost of her later rehabilitation care at a nursing home. Mr. Goldstein said the couple were appealing $13,000 they paid for her nursing home care.” This information can directly correlate to my Hospital data set. Although I do not have clear variables outlining outpatients I do have their reason for disposition and expected source of payment. Perhaps by analyzing these two variables together, along with length of stay and overall cost, I can analyze the diverse charges encountered by distinct insurance companies along with diverse reasons for leaving the hospital. This track may be slightly diverse than the article, however the article does emphasize the importance of knowing the charges a patient is encountering and acknowledging the fees that go along with medical procedures. I feel as though disposition and payer, although un-similar to patient status will still illustrate robust comparative measures and the broad charges correlated with them (Carrns).

Recently there have been proposals made by Congress to help Medicare patients, creating their own definition for inpatient and outpatient. Technically the adopted rule is that after "two midnights" a patient will be considered inpatient. However there have been extreme concerns raised by hospitals and doctors for they believe it is a flawed system. The Hospital data set that I possess considers 1 day overnight as an inpatient. In addition, doctors and case mangers recommend patients ask for their status, and request to have it altered while still in the hospital, if necessary.

Furthermore, due to this articles specific concerns about Medicare and Medicaid patient statuses, I will be applying careful attention to Medicare and Medicaid patients, their length of stay and total cost, in comparison to other forms of expected insurance payments. The Hospital data was recorded during the time period of this article, hence we can either see the significant correction to the current topic surrounding differing charges to patients. Summary statistics analyzing the mean, median, mode, and etc. of the data, ANOVA analyzing the significance of variables, and graphical representations, such as bar charts, will all be utilized in my data analysis in both SAS Miner and SAS Enterprise Guide.

"Hospital Billing Varies Wildly, Government Data Shows" was published in 2013 also compares prices of hospital visits. However rather than relating inpatient to outpatient they are correlating prices between locations. Both articles are looking specifically at Medicare and Medicaid patients. The article states "Data being released for the first time by the government on Wednesday shows that hospitals charge Medicare wildly differing amounts- sometimes 10 to 20 times what Medicare typically reimburses- for the same procedure, raising about how hospitals determine prices and why they differ so widely." Medicare does not pay the hospital stay bill, but instead utilizes a standardized system reimbursing patients a standard amount for the charges they incurred. Representatives from hospitals, supporting the hospitals actions, claim that the discrepancy in charges is due to the health of patients prior to surgery and their overall length of stay in the hospital, not because the hospital is charging more than average for a particular service or procedure (Meier, Mcginty, and Creswell).

There is extremely limited public knowledge about the charges of diverse doctors and hospitals. Although Congress and the government are making a robust attempt, there is no true rational base for the charges of services at distinct hospitals. In general, non-profit hospitals tend to charge significantly lower service fees than profit making hospitals. In addition there are numerous specific examples illustrating the price discrepancy. For example, to implant a pacemaker in Livingston, N.J. is cost an average of $70,712 while in Rahway, N.J. patients are charges approximately $101,945 a 44% increase. It is evident that hospital charges vary immensely amongst location (Meier, Mcginty, and Creswell).

The Hospital data set that I possess, although it has over 1.5 million records, only contains patients from the state of Rhode Island. Based on the previous New Jersey example clearly charges can differ drastically between states, not just regionally. In the Hospital data set, there are fourteen distinct hospitals throughout the state of Rhode Island that shared their patient information. I would be extremely curious to see the differing charges for identical procedures between patients. Although it will be impossible to set every variable such as age, length of stay, and admission date equal, I will try my best to compare extremely similar patients, under either Medicare or Medicaid to examine whether or not there is a price discrepancy. If all other variables are equal, or just about equal, I will be able to compare identical procedures provided at distinct providers and relay back a conclusion on whether or not there is an apparent difference in hospital charges. I believe running summary statistics, and ANOVA on the variables total cost and provider in the SAS Software would be extremely beneficial in determining a final conclusion.

The final article from The New York Times that I will be utilizing for my research is “When Hospitals Buy Doctors’ Offices, and Patient Fees Soar.” Comparable to the previous two articles, this publication speaks about the price discrepancy for Medicare patients when a hospital buys out a doctors’ office. Medicare patients commonly pay a 20% copay so a change, or increase, in the charge drastically impacts their out of pocket expense. The article affirms, “Cardiologists are not the only doctors who have been migrating toward hospital practice. In the last few years, there have been increases in the number of doctors working for hospitals across the specialties. And spreads between fees for office services exist in an array of medical services, down to the basic office visit.” Administration is proposing a flat charge to Medicare patients based on the service or procedure provided. Obviously, there is extreme backlash from hospitals, for private-practice doctors charge significantly lower fees for services and procedures, hence hospitals would incur a higher charge for taking in Medicare patients if an average flat fee was charged (Sanger-katz).

This article reaffirms the previous two articles from The New York Times. Medicare and Medicaid patients are being charges at a disproportional amount. Unfortunately I do not have data to compare the cost of procedures from the Rhode Island Hospitals to that of the private-practice doctors in Rhode Island. However these articles have spurred a question of inquiry of my own. As stated previously, I would like to see if there is a difference between services based on different expected sources of payment. For example, would a male age 68 with Blue Cross Blue Shield needing a pacemaker be charges differently than a comparable male age 65 with Medicare? As we have learned from the previous articles other factors in the data set, such as length of stay, previous health conditions, and location would impact the overall cost as well. It will be extremely interesting to compare identical patients and analyze the difference, or similarity, in their average cost per day.

Healthcare Cost and Utilization Project Statistical Briefs

The final paper I utilized focuses on quantitative values specifying the exact number of health care costs in the years 2003 and 2011. In addition it predicts the specific number, of both length of stay and total healthcare charges, for the year 2013. "Trends and Projections in Inpatient Hospital Costs and Utilization, 2003-2013" is a statistical brief published by HCUP. HCUP is also known as the Healthcare Cost and Utilization Project was established in 1988, they gather statistical data from State data, the Federal government, and private organizations creating databases about cost and quality of health services. I am focusing primarily on statistical brief umber 175. Average health care costs and lengths of stay for patients are disclosed for the years 2003 and 2011, and projected for the year 2013. The average cost per day in 2003, based on 2013 dollars, is $1,915.79. Moreover, the average cost per day predicted for 2013 is $2,391.30. In addition the costs and lengths of stay are specified for diverse hospital services such as surgery and medicine. It transparently states that surgical services possess and increased cost per day in comparison to maternal costs. Moreover, the article emphasizes information on the discharge of patients across the United States. Service is an immensely large portion of my project, for it influences the cost per day of a patient significantly. Although I will need to examine Length of Stay and Total Cost separately, this article is an amazing scholarly research to compare my statistical discoveries to (Barrett, Steiner, and Weiss).

Furthermore, the HCUP publishes other statistical briefs that although I do not possess a primary focus on, could be utilized in my evaluation of payer. “Costs for Hospital Stays in the United States, 2012” contains similar information to the previous article, however, there is one significant difference. The article is also known as statistical brief number 181, and contains information on patient ages and primary payers. It contains the data from 2003, 2008, and 2012 articulating the mean hospital costs per stay and proportion of people in the United States that each specified insurer covers. Furthermore, the article pronounces that Medicare costs are commonly near $12,200, private insurance costs are approximately $9,700 and Medicaid incur the lowest cost at $8,100. This will directly correlate to the Payer independent variable in my data set. Both the New York Times and HCUP articles will undoubtedly assist me in affirming my statistical uncovering’s and guide me in directions for further analysis (Elixhauser, Levit, and Moore).

HYPOTHESES/ASSUMPTIONS

Hypotheses

To commence, my null and alternative hypotheses possessed an immensely broad outlook on the data. The null hypothesis that I proposed stated that the cost per day means of the input variables are equivalent, or as a statistician would inscribe $H_{o:}\ u_1 = u_2 = u_3 = u_4 = u_5$. This is the null hypothesis for all statistical experiments. The alternative hypothesis pronounces that at least one of the cost per day means is not equal. Although the null and alternative hypothesis retain a vast position on the data, based on the background research I can undoubtedly propose assumptions about the data set.

Assumptions

There are a variety of independent variables that I believe may heavily influence the average cost per day at a hospital. Based on the background research, Service, Age, Disposition, and Payer have gradually gained prominence in influencing the charges attained at a hospital. Service encompasses the diverse motives as to why patients visit the hospital, for example surgeries, pregnancies, and medicine. Moreover, one would expect age to cause a distinction amongst explanations for health visits. Furthermore, disposition is an extremely expansive category ranging from death, to leaving without consent of the examiner. Finally, payer dictates whom whether it is the government, or private insurance companies, or the patient will be incurring the cost per day charge. All of these fields are likely to have a prominent significance on the average cost per day of a hospital stay.

In addition to assumptions based on the background research, I arrived at this project with personal biases. Whether it is on the news, online via articles, or through word of mouth it is commonly stated and readily acknowledged that men encounter health issues, and die at a much younger age, when in comparison with women. For this reason, I presumed that there would be a stark difference between the cost per day of women and men. Although personal assumptions and biases may retain validity, it is important to remember, the overarching null hypothesis to prevent discrimination in the data analysis.

METHODS

Introduction

The project that I am completing is an immensely analytical endeavor and requires software systems that can analyze substantial amounts of data. Very few statistical tools possess the ability to evaluate a 1.5 million row data set. Due to the immense amount of data I will be utilizing SAS Enterprise Guide, SAS Enterprise Miner and Tableau. SAS possesses a great number of extremely beneficial advantages. The gentle learning curve, expedient results, and proficiency in big analytics and large data. Moreover, the capabilities in visuals and predictive modeling enhance the positive rationalizations for my utilization of this software. Moreover, both my advisor and I partook in extensive training in SAS software, hence it was the logical selection.

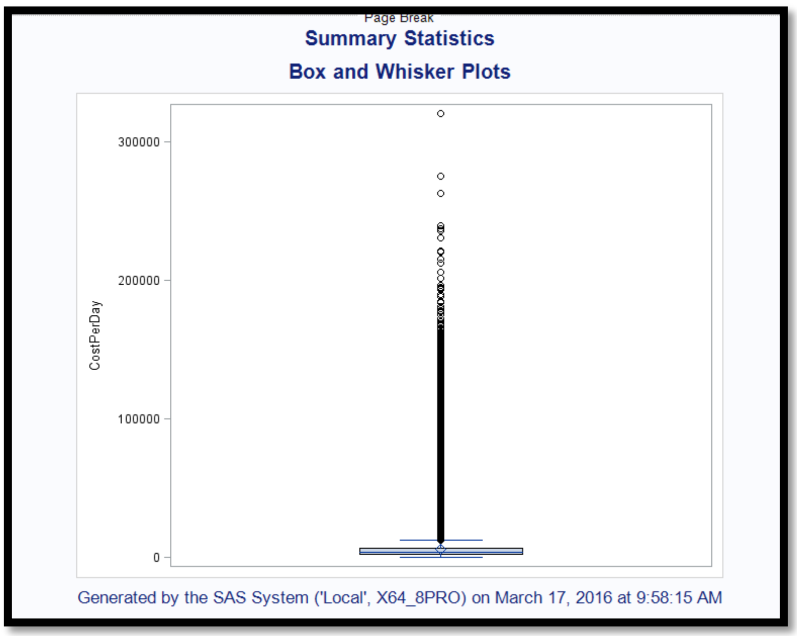
Data Cleaning

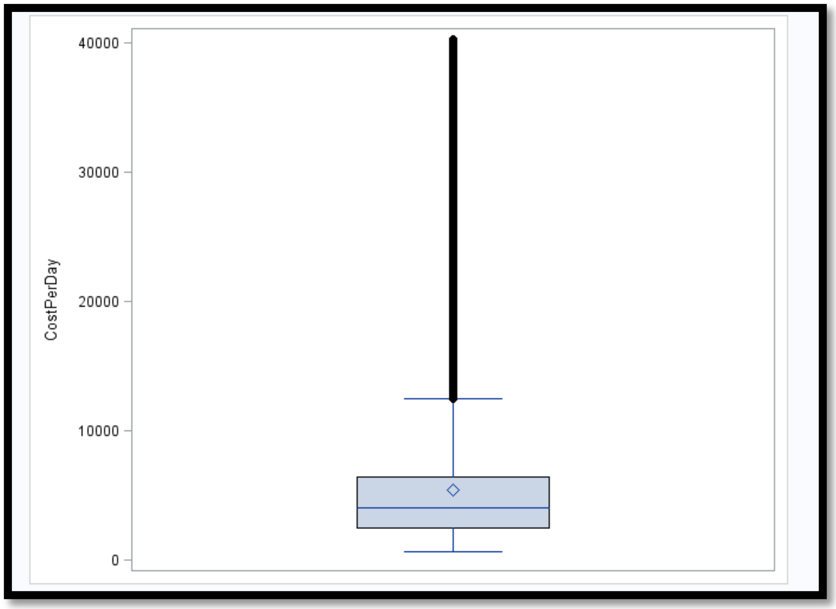
Prior to any analysis data cleaning must be performed to normalize the data. I commenced by removing extraneous variables based on the results from the summary statistics and one-way frequencies utilized in SAS Enterprise Guide. Summary statistics are performed on quantitative variables and one-way frequencies are carried out on qualitative variables. For example, I removed the row where Sex was equivalent to the number 9. Although this may represent transgender, or that the patient desires not to disclose their gender, this cannot be confirmed hence it is obligatory that it is removed from the data set. Moreover, there were an extensive number of negative LOS, length of stays, TOT, total costs, and cases where the pre-operation days exceeded the total length of stay. These specific rows were removed from the data set for the information was unquestionably incorrect.

In addition, the columns Cost Per Day and MDC, Major Diagnostic Categories were computed through query builder based on advanced expression and recoded column. Cost per Day was created by dividing the Total Cost by the total Length of Stay, after the extraneous values were removed. The major diagnostic categories were creating by replacing ranges based on the DRG's, or diagnosis related groups. There are approximately twenty-seven grouping that are distributed online from both medical and government data sources. For example, circulatory system, transplants, and digestive system are merely a few of the categories the nearly one-thousand DRG's can be grouped into. Finally, I created the column Cost per Day categories based on the four quartiles based on the cost per day summary statistics. The variables utilized in analysis are illustrated below.

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
AgeCat	Input	Ordinal	No		No	.	.
CostPerDay	Target	Interval	No		No	.	.
MDC	Input	Nominal	No		No	.	.
admtype	Input	Nominal	No		No	.	.
age	Rejected	Interval	No		No	.	.
asource	Input	Nominal	No		No	.	.
dispub92	Input	Nominal	No		No	.	.
drg	Rejected	Nominal	No		No	.	.
los	Rejected	Interval	No		No	.	.
payer	Input	Nominal	No		No	.	.
preopday	Input	Interval	No		No	.	.
provider	Input	Nominal	No		No	.	.
raceethn	Input	Nominal	No		No	.	.
service	Input	Nominal	No		No	.	.
sex	Input	Nominal	No		No	.	.
tot	Rejected	Interval	No		No	.	.
yod	Input	Interval	No		No	.	.

Removing outliers, filtering out unnecessary columns, and altering the column formats were also tasks I encountered during the data cleaning process. I removed the outliers above the 99^{th} and below the 1^{st} percentile to normalize the data. I also temporarily removed columns from the data set, for my advisor and I determined that they were not pertinent to analysis, and other columns were more beneficial in determining the cost per day at a hospital. A box and whisker plot of the cost per day, before and after outliers were filtered out is following. Finally, in order to transport the data in SAS Enterprise Miner there could not spaces in the columns names, and they needed to be formatted correctly as either character to numeric variables. One of the problems that arose was character columns being formatted at $ColumnName. For example, Sex was characterized as $Sex when it ought to be formatted as $Char8. I manually modified each column to be formatted correctly prior to entering the data in SAS Enterprise Miner.





SAS Enterprise Miner

The primary drive for my methods section was upon the basis of my null and alternative hypotheses. I commenced my evaluation by transporting the SAS data file into SAS Enterprise Miner. Miner requires slightly more extensive start-up work than SAS Enterprise Guide. A new project is produced, with a file name and where in your c-drive it ought to be saved, a library, where the actual SAS file is stored for analysis, must be established, and a diagram is created, where the predictive modeling is completed is generated. Data partitions, splitting the data between training and validation, impute nodes, reassigning numbers to missing values, and a stat explore of variables, to examine the distribution of the various variables, are dragged into the model prior to any predictive analysis. To be more specific, there was missing values in nearly all of the independent variables in the data set. In the impute node I had chosen to leave the missing data in the model as a period, for it does not overtake the significant data points. In addition, I solely changed the pre-operation day variable to zero instead of missing, for most hospital services, other than surgery, do not require pre- operation days.

For the predictive analysis portion, decision trees, regressions and neural networks were employed to test for significance amongst the independent variables in predicting the mean cost per day at a hospital. Three decision trees were evaluated, a chi square tree, and entropy tree, and a gini tree. A decision tree categorizes the input variable by significance of cost per day means, creating a path guiding the user in route of most and least meaningful units. Each possess a distinct number of maximum branches, or the number of splits in a tree, maximum depths, or the number of nodes after a split, and minimum categorical units, or the number of data points necessary to create a new node, to alter the results of the decision trees. The entropy tree possessed the lowest average squared error, hence it fit the model in determining the cost per day category the most accurately in comparison to the other decision trees. The entropy tree retained properties of three maximum branches, a maximum depth of four, and a minimum categorical of five. The regressions and neural network possessed average squared errors even lower than that of the decision trees.

Unfortunately neural networks are exceedingly complicated to explain, and although they may fit a specific model perfectly, they are not utilized in relaying results for their lack of clarity and un-ability to fit scored data. For this reason, I will be predominantly focusing on the regression. A regression creates a formula based upon the dependent and independent variables in the model. Hence this regression equation focuses on cost per day, the y, and the input variables as the X's in order of significance. In the properties panel, the selection model needs to be changed to stepwise, the selection criterion altered to validation error, and the use selection defaults as yes. Finally a model comparison was brought into the diagram. A model comparison compares the average squared errors for each predictive modeling technique run, to determine the best-fitting case for the model, which in this case, ignoring the neural network, is the stepwise regression. Average squared error compares the difference between the predicted and actual value. As expected, the 1.5 million row data set, produced a greater average square error than ordinary.

An attempt to do a cluster analysis was also completed in SAS Enterprise Miner. However, the results were detrimentally limiting pulling all of the data into two vast clusters. The cluster analysis, although useful to some analyses is not beneficial in the Hospital Data hence I altered my plan of action moving forward to do analysis in SAS Enterprise Guide with one-way ANOVA's and Linear Models.

<u>SAS Enterprise Guide</u>

To commence I ran five distinct one-way ANOVA's in SAS Enterprise Guide. Cost per day by Service, Provider, Payer, Year of Discharge, and Major Diagnostic Category were evaluated. One-way ANOVA's test whether there is a difference in the mean cost per day within the variable. The Bonferroni t test and Tukey's studentized range tests for significance of means amongst the individual pairings within the independent variable. For example, in the YOD analysis 2003 would be compared to each year from 2004-2013 separately testing for significance. Moreover, a p-value of $<.05$ represents significance in the overall model and the symbol "***" conveys significance amongst the individual tests, for their confidence intervals do not overlap.

In addition to the one-way ANOVA tests, I ran two linear models testing for significance against the cost per day mean. Linear models creates a formula, similar to regressions, however they utilize one level amongst each input variable as a base for their estimates. Cost per day by YOD, Provider, MDC, and grouped by Service was the primary linear model ran, for these variables were immensely significant in both the decision trees and regression performed in SAS Enterprise Miner. Moreover, a linear models was then performed on cost per day by Age Categories, Sex and Race/ Ethnicity to also test for significance. Finally, in this specific working session, a rapid predictive modeler was run in SAS Enterprise Guide and opened in SAS Enterprise Miner.

A rapid predictive modeler takes all of the independent variables and the dependent variable in a data set and provides, what SAS believes, is the best model to fit the data. The process flow executed in Miner is quite similar to the process I previously described, however there are a few notable differences. Prior to exercising a data partition, SAS decided to take a random sample of the data set, and utilize this data instead of the full 1.5 million rows. This will undoubtedly cause distinctions between the results in the future. The rapid predictive modeler chose to run an impute node, a stat explore node, and a decision tree which is extremely comparable to the steps that I took previously. However, as I had made the decision to run multiple variations of the decision tree node, SAS followed the path of executing multiple regressions. A stepwise, main effects, and forwards regression were completed. A model comparison was utilized to compare the average squared errors of the predictive modeling techniques, and this time, the decision tree possessed the most accurate model in predicting cost per day. The results of the rapid predictive modeler will later be compared to my own executions in SAS Enterprise Miner.

Visuals SAS Enterprise Miner

The final portion of my analysis of cost per day was creating visual representations of the data. Although quantitative data is necessary to relay conclusions, illustrations are simplistic to understand and assist with the delivery to non-statisticians. Both SAS Enterprise Guide and Tableau are software systems that convey distinguished and well-defined graphics, hence I utilized these systems in creating the visuals to convey significance amongst the cost per day column.

Visuals in SAS Enterprise Guide include but are not limited to scatter plot, line charts, pie charts, bar charts, bubble pots and title charts. Pie charts can be stacked or grouped by independent variables. Moreover, bar charts can utilized in a similar manner and be stacked, segmented, or color coded. Bubble plots can be possess one independent and one dependent variable, and a frequency or count to represent the size of the bubble. Bubble plots, comparable to the other charts, can be grouped by another variable. Title charts are identical to bubble plots, however their final product is a rectangular shape with diversely sized rectangles amongst the outline. Finally, scatter plots and line charts create a straightforward visual of the cost per day mean either by category or over time. The dependent variable in the visual representations was cost per day and the independent variables are both quantitative and qualitative such as Payer, YOD, MDC, Service, and Age.

Visuals Tableau

Tableau executes immensely similar visuals to those of SAS Enterprise Guide however, the colors are more prominent, and there are a wider variety of graphs available. Although, one problem with Tableau is the in-ability to handle vaster data sets. For this reason, it was obligatory that I take a 10,000 record subset of the SAS file and transport it into Tableau for evaluation. The illustrations undoubtedly assist in the understanding and interpretation of significant results for the cost per day.

Conclusion

SAS Enterprise Miner, SAS Enterprise Guide and Tableau are immensely useful and broadly utilized tools for data analysis. These tools possess a vast number of capabilities to analyze data, a few that I was even unaware of prior to this process. The chosen dependent variable of cost per day, allowed itself for significance amongst the considerable amount of independent or input variables. This statistical software systems employed are extremely valuable to learn in depth, for they are utilized throughout a countless number of fields in the workforce. Not merely did they assist in the evaluation, but contributed to new pathways of investigation throughout the process, enhancing the overall results.

RESULTS

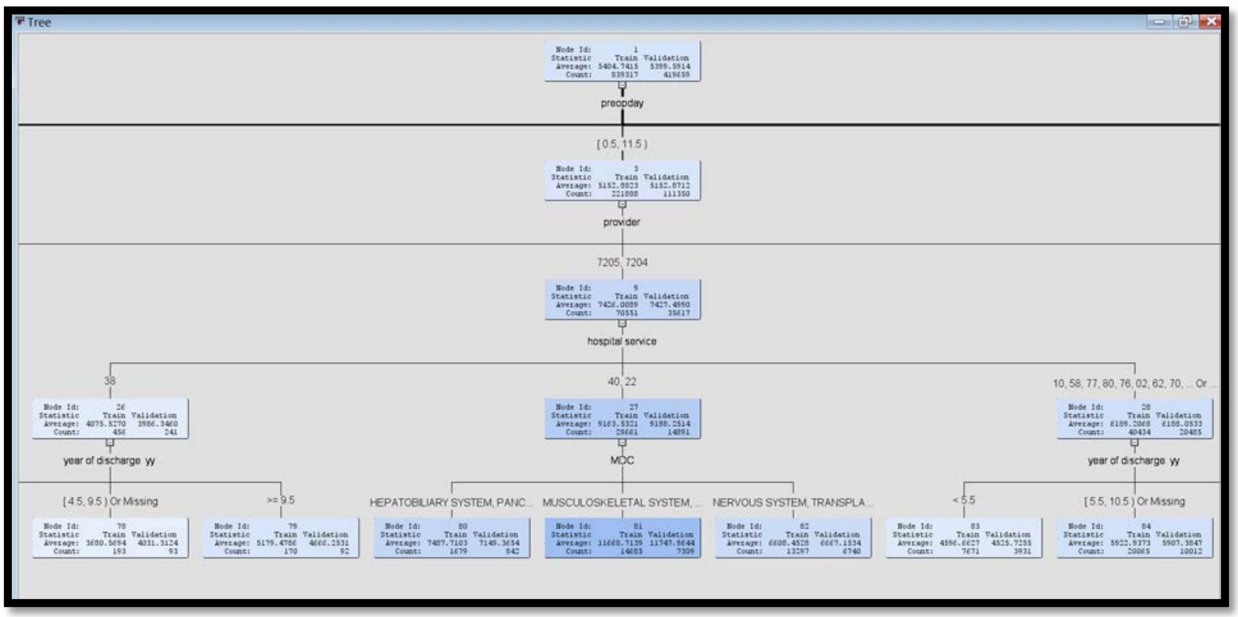
SAS Enterprise Miner

The Entropy decision tree possessed the lowest average squared error, hence those are the results that I will be delivering. The average squared error is 13,640,726. The first split in the decision tree was due to pre-operation days. To commence the validation, for the cost per day at the hospital, was $5399.59. If the pre-operation days was greater than or equal to 11.5 days than the validation, for cost per day, surprisingly decreased to $3673.66. However, if the pre-operation days was less than .5 days the cost per day increased substantially to $7851.00. From this point onward, I will commence on explaining the path of "greatest" cost per day followed by following the "least" cost per day journey.

After a pre-operation day of less than .5 days, the decision tree is separated based upon MDC, or Major Diagnostic Category. The categories Hepatobiliary system and Pancreatic system possessed an elevated cost per day of $10,323.02. The final two divergences, in this portion of the decision tree, are due to admission source and hospital service. Admission sources 4, 6, 2, and D also known as Trans-hospital, Trans-Health care, Clinic referral, and Extramural Birth retained an average validation cost per day of $15,063.73. Finally, the most extreme cost per day's final determinant was Hospital service. Services 40, or surgery, acquired an immensely extreme cost per day of $17,928.36.

Next, I will be describing the "least" cost per day path at a hospital. After a pre-operation day of greater than or equal to 11.5 days the next divide is based upon Provider. Hospitals such as Butler, Bradley, and Rehab of RI possess a decreased cost per day of approximately $2,703.24. From there, there are two splits by hospital service. The first is based on services

80, 76, 02, 62, 70 and more. Newborn, OB-Not Delivered, Pediatrics, Urology, and Gynecology are the hospital services relating directly to a cost per day reduction to $1,652.08. The final split, as I stated is also hospital service, specifically hospital service 80, or newborns. The final, "least" cost per day, based on this direction, is $940.28. A portion of the overall decision tree can is illustrated below.



In addition to the decision tree a stepwise regression was performed on the dependent variable cost per day. The average squared error for the regression, is 13,624,062, which is lower than the decision tree representing a better fit of the data. The first step, is the intercept, followed by the independent variable Service. Followed shortly by the distinct input variables YOD, or year of discharge, Provider, MDC, or Major Diagnostic Category, Admission Type, Disposition, Pre-Op days, Payer, Admission source, and Age Categories. All of the variables were significant in the regression with a p-value of <.0001. Based on the model comparison seen following, either predictive model, the decision tree or the regression, could be utilized to determine the cost per day at a hospital of an inpatient.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Average Squared Error
Y	Neural2	Neural2	Neural N...	CostPer...		13251133	13240468
	Reg	Reg	Regressi...	CostPer...		13550591	13624062
	Tree2	Tree2	Decision ...	CostPer...		13587177	13640736
	Neural	Neural	Neural N...	CostPer...		14306013	14384777
	Tree	Tree	Decision ...	CostPer...		15968244	15968488
	Tree3	Tree3	Decision ...	CostPer...		15968244	15968488

<u>SAS Enterprise Guide</u>

One-way ANOVA's and Linear Models were completed in SAS Enterprise Guide on the dependent variable cost per day to test for significance amongst the means. I tested Service, Provider, YOD, or year of discharge, and MDC, or major diagnostic category against the cost per day, independently, comparing the cost per day means of each input variable. All of the one-way ANOVA's run were significant with a p-value of <.0001. The r-squared measure in ANOVA illustrates the amount of variation in the cost per day that can be explained by the independent variable in the model. In general terms, it conveys whether or not the test is an overall good fit for the model. Although it is not exceptionally high, the r-squared related to the Service one-way ANOVA was the best model I ran with an r-squared value of 21.27%. The r-squared value with Provider is 12.09%, YOD is 5.74%, MDC is 17.20% and Payer is 1.78%.

In addition to solely the one-way ANOVA, post-hoc tests can be explored if the model is significant. For example, I utilized the Tukey Studentized Range (HSD) test for comparing the means of the cost per day. This test exercised utilizes pairings, or levels, within the distinct independent variables and testing their significance individually. To be more specific, for the one-way ANOVA run against payer there are a wide variety of methods of payment. 0 in the study represents Medicare. Based on the Tukey test, Medicare is significant when compared to all of the other methods of payments expect for Unknown and Neighborhood Health Plan of RI. Medicare is significant against Medicaid, Blue Cross, Workers compensation and others. Another example would be the comparison by Provider. Provider 7216, was significant when compared to all of the hospitals except 7212. 7216 is Butler and 7212 is Rehab of RI hence their non-significance is expected for they provide similar services. This process of comparing the independent variables at the lowest level was completed for each of the one-way comparisons. Specifically, the one-way ANOVA for payer and cost per day, along with the Tukey test for Medicare are conveyed below.

One-Way Analysis of Variance Results

The ANOVA Procedure

Dependent Variable: CostPerDay

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	601780108784	42984293485	1812.93	<.0001
Error	1.4E6	3.3166358E13	23709782.404		
Corrected Total	1.4E6	3.3768138E13			

R-Square	Coeff Var	Root MSE	CostPerDay Mean
0.017821	90.13304	4869.269	5402.313

Source	DF	Anova SS	Mean Square	F Value	Pr > F
payer	14	601780108784	42984293485	1812.93	<.0001

Tukey's Studentized Range (HSD) Test for CostPerDay

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	1398847
Error Mean Square	23709782
Critical Value of Studentized Range	4.79593

0 - 4	-3722.527	-3945.525	-3499.529	***
0 - 8	-984.148	-1197.686	-770.610	***
0 - D	-663.563	-730.402	-596.724	***
0 - X	-309.169	-443.893	-174.446	***
0 - 7	-91.847	-163.532	-20.161	***
0 - H	-23.651	-258.999	211.696	
0 - 6	348.719	293.622	403.817	***
0 - E	369.694	285.358	454.031	***
0 - 5	452.197	413.212	491.182	***
0 - Z	971.228	-88.303	2030.760	
0 - Y	1182.494	188.290	2176.697	***
0 - G	1387.126	1342.067	1432.185	***
0 - B	1409.451	1255.053	1563.848	***
0 - 1	1635.736	1573.874	1697.599	***

Finally I ran two distinct linear models. The first linear model run possessed the dependent variable cost per day and independent variables service, provider, year of discharge, and major diagnostic category. Not merely was the model significant with a p-value of <.0001 but the type three sum of squares was significant for each independent variable. The r-squared for this model was .378. Hence 37.8% of the variation in cost per day can be explained by the independent variables, which although may not be a perfect fit for the model, it not a bad fit of the data. As explained previously, each independent variable retains a base in the creation of a formula for the model. The bases are service 98, or rehabilitation, provider 7216, or Butler, a 2013 year of discharge and transplants as the major diagnostic category. When compared to service 98, all of the other services, such as service 10, medicine, or service 40, surgeries were significant with p-values of <.0001. Moreover, all of the providers, for example 7205, Rhode Island Hospital, and 7214, Women and Infants Hospital, are significant when compared to the base, Butler Hospital. In addition, all of the years of discharge were significant with a p-value 0f <.0001 when paralleled to 2013. Finally, the major diagnostic category circulatory system, was the only major diagnostic category not significant when equated to the MDC transplants. Although this linear model performed provided significant results that was also a good fit for the model, I also ran a linear model illustrating significance of diverse input variables, sex, age categories, and race and ethnicity. The linear model output is illustrated below.

Linear Models

The GLM Procedure

Dependent Variable: CostPerDay

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	65	1.2759677E13	196302719122	13073.9	<.0001
Error	1.4E6	2.0988002E13	15014853.47		
Corrected Total	1.4E6	3.3747679E13			

R-Square	Coeff Var	Root MSE	CostPerDay Mean
0.378090	71.71845	3874.900	5402.934

Source	DF	Type I SS	Mean Square	F Value	Pr > F
service	16	7.1814011E12	448837566938	29892.9	<.0001
provider	13	2.0363498E12	156642293504	10432.5	<.0001
yod	10	1.9257661E12	192576613523	12825.7	<.0001
MDC	26	1.6161597E12	62159989274	4139.90	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
service	16	1.9961596E12	124759975865	8309.10	<.0001
provider	13	1.589839E12	122295308787	8144.96	<.0001
yod	10	1.770664E12	177066404815	11792.7	<.0001
MDC	26	1.6161597E12	62159989274	4139.90	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	2238.990423	B	89.4550306	25.03	<.0001
service 00	3572.763415	B	94.1458211	37.95	<.0001
service 02	3395.019980	B	81.6431950	41.58	<.0001
service 10	3943.423560	B	76.9562375	51.24	<.0001
service 22	5886.261882	B	81.4006812	72.31	<.0001
service 38	2451.227954	B	79.1045157	30.99	<.0001
service 40	6977.074877	B	76.8679434	90.77	<.0001
service 48	4993.411146	B	98.5335505	50.68	<.0001
service 50	4406.572395	B	110.4857160	39.88	<.0001
service 54	5114.935779	B	128.3676927	39.85	<.0001
service 58	4606.559568	B	76.8493168	59.94	<.0001
service 62	6752.008949	B	81.5187258	82.83	<.0001
service 70	6848.184777	B	84.3853318	81.15	<.0001
service 75	6162.398598	B	246.1300798	25.04	<.0001
service 76	4504.018893	B	92.6318096	48.62	<.0001
service 77	4797.193111	B	79.9239885	60.02	<.0001
service 80	2491.135458	B	80.1320943	31.09	<.0001
service 98	0.000000	B	.	.	.
provider 7201	-654.899171	B	26.8987713	-24.35	<.0001
provider 7202	-157.750674	B	24.4324130	-6.46	<.0001
provider 7203	135.257821	B	26.8917884	5.03	<.0001
provider 7204	2789.976915	B	24.9649006	111.76	<.0001
provider 7205	2272.215888	B	23.2707260	97.64	<.0001
provider 7206	-419.652445	B	25.2058676	-16.65	<.0001
provider 7209	-79.396653	B	27.8387668	-2.85	0.0043
provider 7210	1138.377644	B	23.8854184	47.66	<.0001
provider 7211	-101.480173	B	29.1259333	-3.48	0.0005
provider 7212	3202.783425	B	96.0170846	33.36	<.0001
provider 7213	447.012362	B	26.0314660	17.17	<.0001
provider 7214	1517.713032	B	26.7880851	56.66	<.0001
provider 7215	-556.285531	B	39.2384065	-14.18	<.0001
provider 7216	0.000000	B	.	.	.

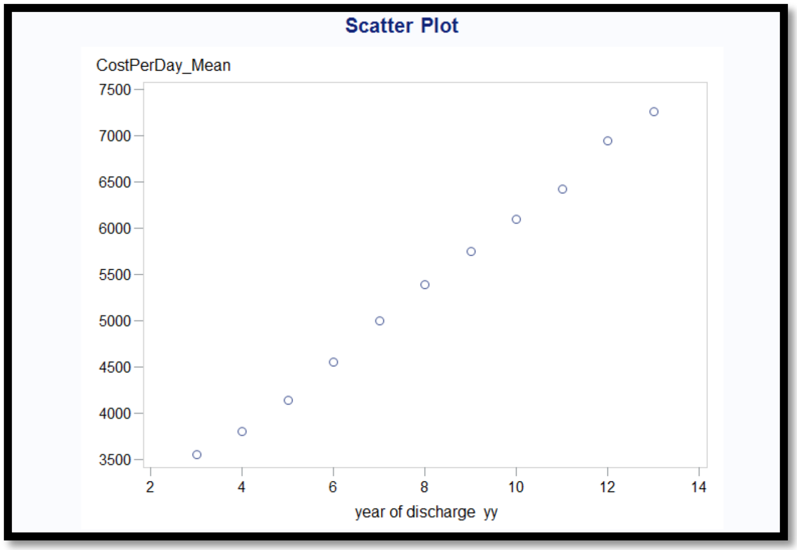
yod 3	-4602.669583	B	17.5423306	-262.38	<.0001
yod 4	-4241.119409	B	17.4123361	-243.57	<.0001
yod 5	-3833.141843	B	17.2950227	-221.63	<.0001
yod 6	-3481.180961	B	17.1107507	-203.45	<.0001
yod 7	-2861.208513	B	16.3164513	-175.36	<.0001
yod 8	-2001.027984	B	15.2251281	-131.43	<.0001
yod 9	-1658.008553	B	15.2531600	-108.70	<.0001
yod 10	-1264.008163	B	15.3331890	-82.44	<.0001
yod 11	-903.419502	B	15.3889722	-58.71	<.0001
yod 12	-350.117818	B	15.4986463	-22.59	<.0001
yod 13	0.000000	B	.	.	.
MDC Circulatory system	-72.701250	B	39.6592735	-1.83	0.0668
MDC Digestive system	-880.393247	B	41.0827719	-21.43	<.0001
MDC Ear, nose, mouth, throat	-612.950724	B	42.4865220	-14.43	<.0001
MDC Eye	1228.622737	B	52.0536198	23.60	<.0001
MDC Hepatobiliary system, Pancreas	286.936627	B	42.3201855	6.78	<.0001
MDC Musculoskeletal system, Connective tissue	3272.439397	B	40.4758831	80.85	<.0001
MDC Nervous system	156.512324	B	40.2255845	3.89	<.0001
MDC Respiratory system	-901.717068	B	40.0810768	-22.50	<.0001
MDC alcohol/drug	-1954.792198	B	49.4993072	-39.49	<.0001
MDC blood, blood forming organs, immunological disorders	-114.134166	B	55.1761078	-2.07	0.0386
MDC burns	-4564.332151	B	165.2042163	-27.63	<.0001
MDC endocrine, nutritional, metabolic system	155.942167	B	46.5182959	3.35	0.0008
MDC extensive procedure	1785.505003	B	91.9471717	19.42	<.0001
MDC factors influencing health status	-1522.133495	B	60.4909567	-25.16	<.0001
MDC femal reproductive system	2985.275482	B	62.2126927	47.98	<.0001
MDC human immunodeficiency virus infection	-633.536051	B	130.7167361	-4.85	<.0001
MDC infectious and parasitic diseases and disorders	-976.692724	B	46.1834676	-21.15	<.0001
MDC injuries, posion and toxic effects of drugs	-396.156917	B	53.5996518	-7.39	<.0001
MDC kidney, uninary tract	-1053.528729	B	45.0300552	-23.40	<.0001
MDC male reproductie system	2025.704708	B	79.7040575	25.42	<.0001
MDC mental diseases and disorders	-1088.751632	B	47.1044112	-23.11	<.0001
MDC multiple sinificant trauma	-254.346598	B	99.0604614	-2.57	0.0102
MDC myeloproliferative diseases and disorders	468.995464	B	63.0208079	7.44	<.0001
MDC newborns, neonates	-2871.831264	B	48.6266811	-59.06	<.0001
MDC pregnancy	-1595.987684	B	48.0055313	-33.25	<.0001
MDC skin, subcutaneous tissue, breast	-1137.736038	B	46.6038848	-24.41	<.0001
MDC transplants	0.000000	B	.	.	.

The second linear model also conveyed significance with a p-value of <.0001. However, this model possessed a substantially lower r-squared value of 4.45%. Solely 4.45% of the variation in cost per day can be explained by the independent variables sex, age categories, and race and ethnicity. Although the values may be significant, the formula, or model, is not a noteworthy fit for the data and may not be utilized in further analysis. The bases for this model are sex 2, or females, age category 65+, and race and ethnicity 9 or unknown race. Compared to their bases, all of the input variable levels excluding the races white and American Indian were significant. The one-way ANOVA's and Linear Models performed in SAS Enterprise Guide are imperative to uncovering new findings and confirming the results discovered through SAS Enterprise Miner.

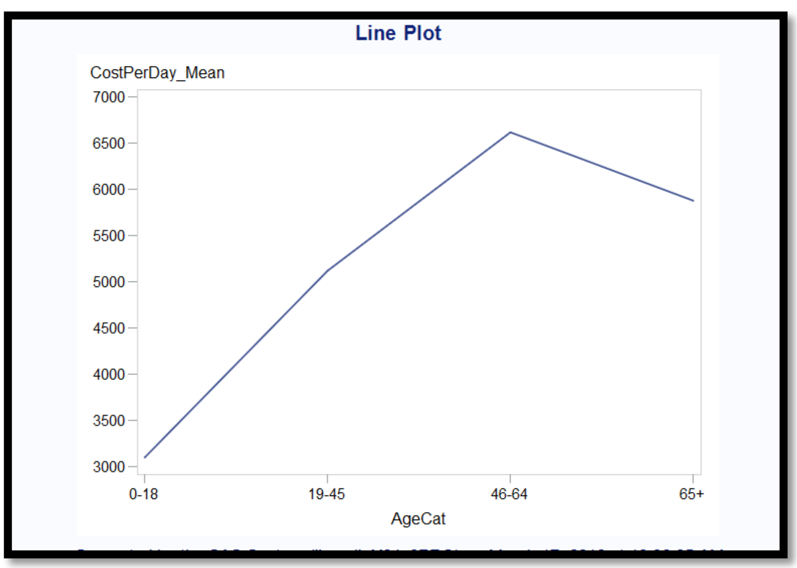
SAS Enterprise Guide Visuals

Visuals are immensely important for clarifying results and illustrating relationships in a manner that is easily comprehendible. I created visuals on both cost per day and cost per day category, by the variables Year of Discharge, Major Diagnostic Category, Service, Provider, Payer, Sex and Age Category. The visuals below are solely developed from SAS Enterprise Guide.

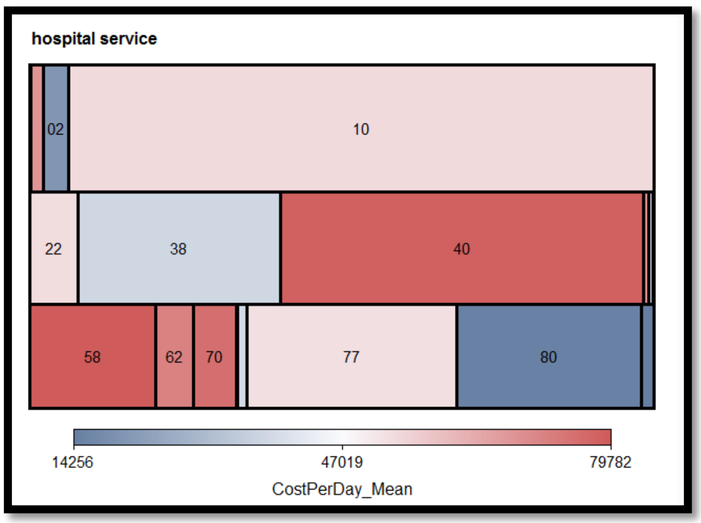
To commence, I created an extremely simplistic scatter plot illustrating the average cost per day over the years 2003 to 2013. There is a positive relationship, illustrating an increase in the cost per day over time. In addition, the relationship even appears to be linear, causing some suspicion as to how hospitals determine their rates and charges. The cost per day has grown drastically from approximately $3,500 in 2003 to about $7,300 in 2013. The increase is most likely due to advances in both medicine and technology, however it may be skewed by information solely comprehendible by hospitals and doctors.



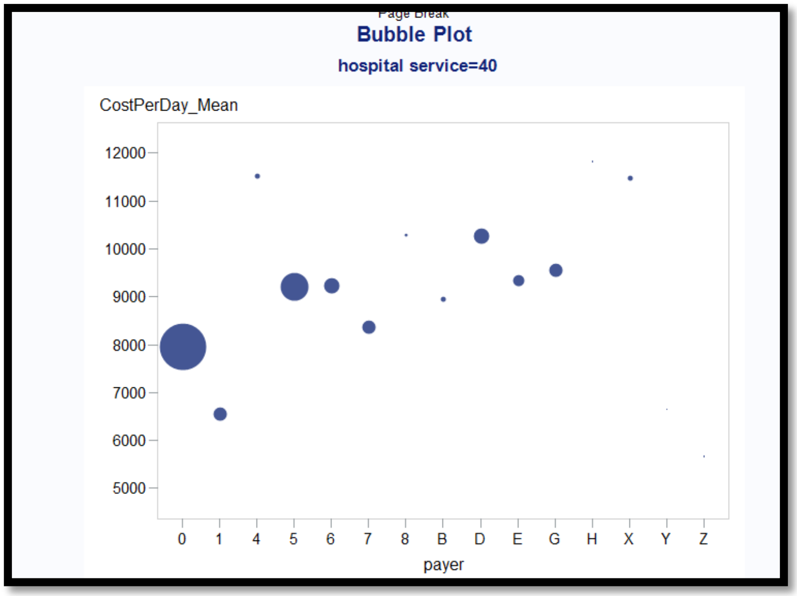
Another simplistic visual was the average cost per day by age category represented through a line plot. The graph reaches a peak at the "46-64" age category with a cost per day mean of approximately $6,700, prior to decreasing to $6,000 at the "65+" category. Upon exploration of this downfall, I uncovered it was most commonly due to the disposition of "expired" causing the decrease in cost per day. This is a surprising discovery followed by a logical conclusion.

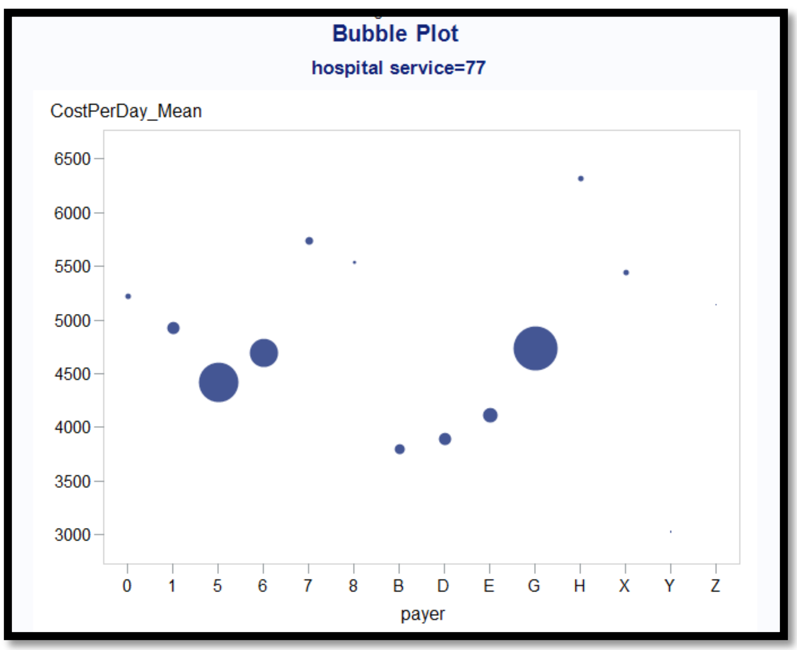


A tile chart illustrated the differences in the cost per day mean by service. The darker reds, representing an elevated cost, represent the services 40, surgery, and 58, orthopedics. The blue, or minimal cost per days are services 02, pediatrics, and 80, newborns. Based on our findings in the decision trees and ANOVA's, this is an expected outcome confirming our previous results.

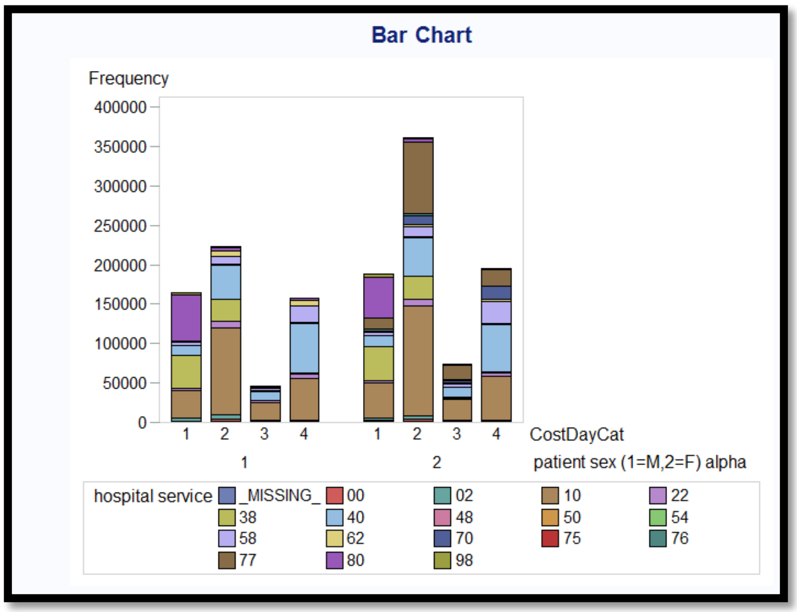


The plots began to become more developed with multiple independent variables. The first plot was a bubble plot of the cost per day mean on the y-axis, payer on the x-axis, grouped by service, with the bubble size representing the frequency of the event. It is evident by the graph, that payer 0 or Medicare represents a large portion of the service surgeries, with an average cost per day mean of $8,000. Medicare covers patients over the age of 65, hence would expect them to need more surgeries due to health declines that arise with age. However, for the service 77, or OB-Delivered, payers 5 and G, or Blue Cross and RIte Care represent the greatest proportion of insurers costing approximately $4,500 per day. Blue Cross is a private insurer, and RIte Care is a Medicaid program available to uninsured pregnant women, families, and children.





The final graph I developed was a bar chart comparing the frequency of services provided by sex grouped by their cost per day category. The colors in the graph represent the distinct services, and the size conveys their frequency. The sole perceptible difference between the males and females, was the observation of service 77, or OB-Delivered, in the bar charts of females. An interesting finding is that services 10, 80, and 40 or Medicine, Newborn and Surgery are apparent at approximately equivalent proportions distributed throughout the cost per day categories for both males and females.

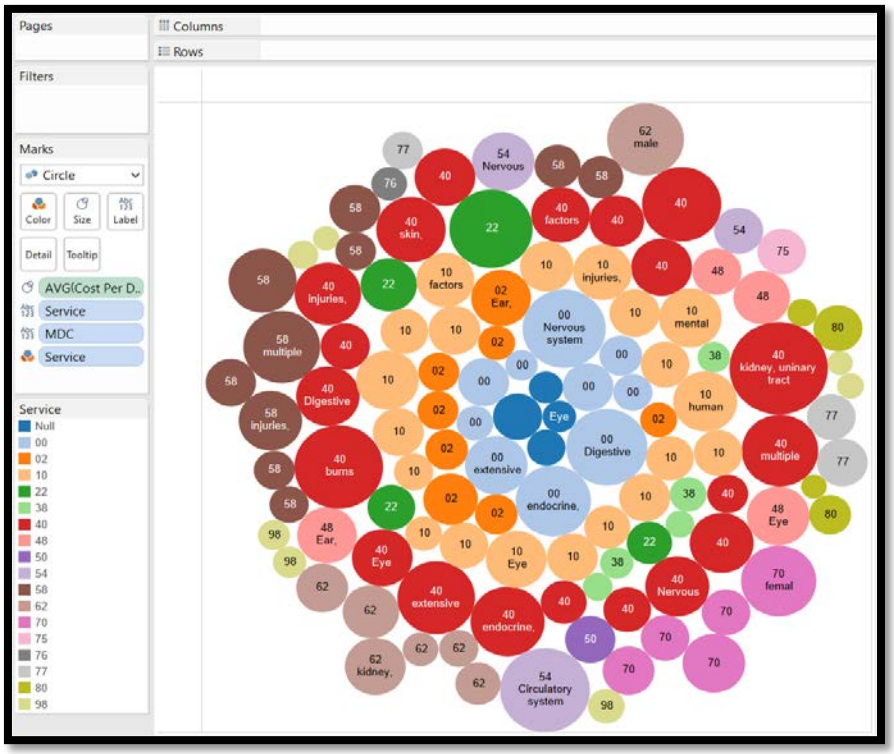


<u>Tableau</u>

Tableau is yet another software system that I utilized for visuals, to convey the significant results, in my project. Tableau, although it possess more visual capabilities and varieties than other systems, cannot handle large amounts of data. In order to attain graphics in Tableau it was necessary that I take a sample of 10,000 records from the 1.5 million data set for visual analysis.

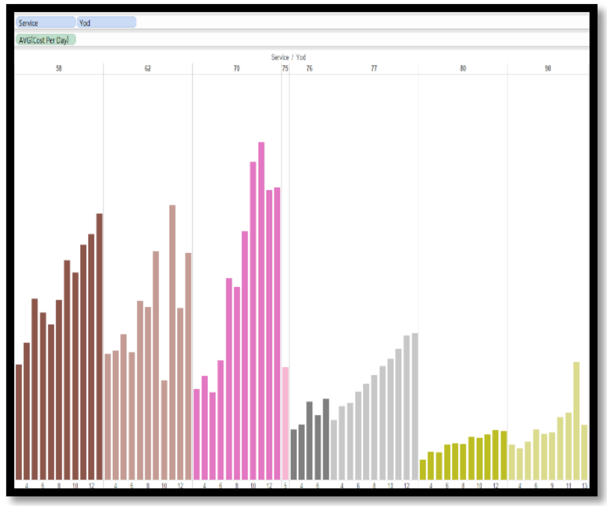
The first illustration that I developed is a stacked bar chart encompassing the variables provider, cost per day, length of stay, and service. The x-axis contains the distinct services, the y-axis conveys the average length of stay, provider is the color and cost per day is the size of the “boxes” on the bar chart. Specifically analyzing service 40, it does not appear that any one provider possesses a drastically longer length of stay when compared to the others, however service 98 or Rehabilitation has a slightly longer length of stay. Moreover, provider 7209, or South County appears to have a slightly lower cost per day while provider 7204, or Miriam retains a slightly more elevated cost per day. Furthermore, it appears that the service orthopedics, or 58, retains the greatest cost per day while newborns, or service 80, have the smallest cost per day.

In addition to the bar chart, I completed a packed bubble graph in Tableau containing the variables cost per day, service, and major diagnostic category. The bubble sizes represent the average cost per day of each service by major diagnostic category. Moreover, the color of the bubbles represent the service, and each bubble is labeled with both the service number and major diagnostic category name. For example, the red bubbles represent surgeries, or service 40. The surgeries possess major diagnostic categories such as burns, kidney, and multiple significant trauma. The packed bubble is a good representation of the overall theme of the data for cost per day for service by major diagnostic category.



The final graph I created in Tableau is a side-by-side bar chart with the variables cost per day, service, and year of discharge. The year of discharge is on the x-axis, the bars are grouped and colored by service, and the y-axis is the average cost per day. Recall the graph I created in SAS Enterprise Guide illustrating the cost per day by year of discharge conveying a nearly linear relationship as time continued. The following graph illustrates a similar result. By service type, the average cost per day tends to increase at a linear rate over time for services such as Medicine, 10, Surgery, 40, and OB-Delivered, 77.


Pages
Columns
Service
Yod
Rows
AVG(Cost Per Day)
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Service / Yod
Null
00
02
10
22
38
40
48
50
54
18K
17K
16K
15K
14K
13K


Service
Yod
AVG(Cost Per Day)
Service / Yod
58
62
70
75
76
77
80
90

CONCLUSIONS

In conclusion, the input variables Service, Year of Discharge, Age Categories, Payer, Major Diagnostic Categories and Provider had an impact on the cost per day, charged, at a hospital. Overall, Service, MDC, and YOD were the most significant in determining the cost per day. Moreover, service 40 or surgeries possess the most elevated cost per day ranging from $17,000 to $40,000 while service 80, or newborns retain a cost per day of merely $950. MDC's are based upon the DRG's or diagnosis related groups. A patient is diagnosed with a DRG upon entry into a hospital, hence it immensely influences the cost per day. As illustrated by the many visualizations in both Tableau and SAS Enterprise Guide, the cost per day has increased at a surprisingly linear rate over time. The other input variables impact the cost per day, however at an extremely insignificant rate when in comparison to Service, MDC, and YOD.

The independent variables such as Payer and Provider do appear to influence the cost per day, however it is commonly due to other factors such as service and MDC. For example, Provider 7214, Women and Infants, possesses a minimal cost per day. However this can be attributed to the fact that Women and Infants specializes in pregnancies and newborns, whom retain the smallest cost per day. Furthermore, the Payer Worker's Compensation retains an elevated cost per day when in comparison to Medicaid programs. However, this may be due to the surgeries that Worker Compensation is covering, for the service surgeries retains the topmost cost per day. It is undoubtedly a common theme, that Service, MDC, and YOD are the underlying factors dramatically influencing the overall cost per day.

To give a specific example, a male, age “65+”, whom needed surgery in the year 2013, under either Worker’s Compensation or Medicare would incur the greatest cost per day surpassing $17,900. Compared to a newborn, born at Women and Infants under either Champus or RIte Care would incur a minimal cost per day of under $1,000.

I would like to emphasize, that the cost per day I refer to throughout the paper is the cost per day the hospital charged. Commonly, the insurance companies do not pay this specific price, but instead pay a proportion of the final cost to the hospital. To be more specific, Blue Cross, Payer number 5, may have an agreement with Rhode Island Hospital, Provider 7205, that they will pay 50% of the total charges incurred. For this reason, it is possible that the doctors will charge more for services exclusively to Blue Cross patients, or that the hospital attempts to have Blue Cross patients in the hospital for a shorter amount of time, or a lesser length of stay, than say a Champus patient. Unfortunately, this information will most likely never be disclosed for these “contracts” are extremely competitive and considered confidential hospital information. For this reason, the cost per day charged is the most accurate information that statisticians and analysts can evaluate to uncover findings, as I have to analyze the cost per day at a hospital.

Finally I would like to promote the utilization of SAS Enterprise Guide, SAS Enterprise Miner, and Tableau. These statistical software systems unquestionably expedited the processes gave me the capabilities to analyze and evaluate a 1.5 million row data set with decision trees, regressions, one-way ANOVA’s, linear models, and visualizations. Through these statistical systems, it is evident that Service, MDC, and YOD possess the greatest influence on the cost per day at a hospital.

REFERENCES

1. Barrett, Marguerite L., Claudia A. Steiner, and Audrey J. Weiss. "Trends and Projections in Inpatient Hospital Costs and Utilization, 2003-2013 - Statistical Brief #175." Healthcare Cost and Utilization Project (HCUP). N.p., June 2014. Web. 03 Mar. 2016.
2. Carrns, Ann. "When a Hospital Stay Is Not a Stay." *The New York Times*. The New York Times, 18 Mar. 2014. Web. 09 May 2015.
3. Elixhauser, Anne, Katharine Levit, and Brian Moore. "Costs for Hospital Stays in the United States, 2012 #181." Healthcare Cost and Utilization Project (HCUP). N.p., Oct. 2014. Web. 03 Mar. 2016.
4. "Healthcare Crisis History." *PBS*. PBS, n.d. Web. 28 Nov. 2015.
5. Meier, Barry, Jo Craven Mcginty, and Julie Creswell. "Hospital Billing Varies Wildly, Government Data Shows." The New York Times. The New York Times, 07 May 2013. Web. 09 May 2015.
6. Sanger-katz, Margot. "When Hospitals Buy Doctors' Offices, and Patient Fees Soar." *The New York Times*. The New York Times, 06 Feb. 2015. Web. 09 May 2015.