FACTORS AFFECTING HOSPITAL FINANCIAL STABILITY

ISM 6137 – Statistical Data Mining
Final Project – Spring 2022

Akib Ali Sardar
Durga Prasad Somarouthu
Sahil Shah
Venkat Krishna Prithvi Kokku

Table of Contents

List of Tables and Figures:

Hospital Cost to Charge Ratio – Analysis Report

Executive Summary

The health-care industry in the United States is the most heavily invested in industry sector. Every year, the US government spends ~$3.65 trillion on federally funded healthcare programs and initiatives. Most prominent among them is Medicare, which is the predominant form of health insurance used by senior citizens. Furthermore, the advent of COVID-19 has brought health-care sector improvement to the forefront of the American agenda. In the past 10 years, 136 rural hospitals have closed due to financial insolvency. This has a major negative impact on communities that depend on those hospitals.

Our dataset consists of cost reports that hospitals from around the country submit to the Centers for Medicare & Medicaid Services (CMS). Our analysis examines data from 18,944 hospitals from across all 50 states and 3 territories compiled over the course of five years (2014-2018).

In our analysis, we consider the impact of key variables on the most commonly used healthcare industry metric to assess financial health - the **Cost to Charge Ratio**. Variables include, but are not limited to, Type of Control, Medicare & Medicaid Ratio, Rural vs Urban designation and Health IT Asset value. We employ Linear Mixed Effects models, controlling for random and fixed effects, to quantify and convey insights.

Key findings include Type of Control being the most significant predictor, a 20+% disparity in Cost to Charge Ratio between Rural vs Urban hospitals, and the favorable effect of investing in Health IT.

This analysis walks the reader through all major steps in the process from data cleaning and feature engineering to model building, their interpretations, and finally recommendations.

`

Problem Definition and Significance

The recent rise in rural hospital closings poses a major threat to communities, especially to the elderly and disabled who rely on those hospitals for medical treatment. 130+ rural hospitals have closed in the past ten years, creating many downstream negative economic effects.

Financial overburden is the primary cause for closings. In 2013, the median hospital lost $82 for each discharge, and only 45% of hospitals were profitable (Bai, G., 2016). The most commonly used metric in the industry to measure hospital financial health is **Cost to Charge Ratio.** Hospitals performing well have Cost to Charge Ratios less than 1.

Since most hospitals, federal health agencies and insurance companies are familiar with Cost to Charge Ratio and its implications, we have chosen it as our target variable. The purpose of this analysis is to identify the major factors that affect Cost to Charge Ratio and provide actionable insights to federal health agencies and hospitals to help improve it.

Prior Literature

We examined numerous existing publications with three primary intentions; First, to better understand the context, breadth and significance of the problem. Second, to understand the variables in our dataset and which are important to consider during modeling. Third, to see how other teams have built and interpreted models with a similar objective.

Eight publications were foundational in developing our understanding of the issue and methodologies for analysis. A brief summary of key takeaways is noted below.

- Hospital solvency is a critical issue in rural areas due to both higher costs and low charges (Balasubramanian, S. et. al, 2016)
- The Type of Control refers to the management type. It can be various types of non profit, government or private control. This has a significant effect on the Cost to Charge Ratio, and thus needs to be considered in our analysis. (Bai, G. et. al, 2016)
- Technological upgrades for IT infrastructure constitute 27% of the total costs, and therefore its effect on Cost to Charge Ratio would be an important factor to consider (Thornton, J.A., 2015)
- The payment from Medicaid and Medicare are set by law rather than through a negotiation process as with private insurers. This would likely have a significant effect on Cost to Charge Ratio which we should evaluate. (Bai, G. et. al, 2015)
- Compared with urban hospitals, rural hospitals generate a larger share of their revenue from Medicare (45%) Thus Medicare's effect on Cost to Charge Ratio may disproportionately affect rural hospitals. (Balasubramanian, S. et. al, 2016)
- OLS models are used in most of the publications for Cost to Charge Ratio as the target variable.

A more detailed summary of each publication, its key points and relevance to our study is included in the appendix.

Data Source, Preparation & Feature Engineering

The original raw dataset is comprised of annual CMS Cost Report data from 2014 – 2018. The concatenated set consists of 129 variables and 31,044 rows. Upon further inspection we found many issues that needed to be addressed before modeling. Cleaning and preparation steps are noted below.

1. Categorical Features:
    - Renamed Provider Types, Types of Control and Rural vs Urban designation to reflect corresponding names from the data dictionary
    - Releveled state factor to FL, Type of Control to Proprietary-Individual and Provider Type to General Short Term
2. Filtered the dataset to include only those rows for which the cost report duration is 364 days
3. Created new Total Days Unknown variable to reflect the number of Total Days not attributed to any federal health insurance title
    - Total Days Unknown = Total Days (V + XVIII + XIX+ Unknown)- Total Days (V + XVIII + XIX)
4. Handled NA values
    - Removed rows with NA for Cost to Charge Ratio, Total Days Variables (XVIII, XIX, Unknown), Total Assets, Total Income, Total Unreimbursed & Uncompensated Care, Total Current and Long-Term Liabilities
    - Imputed value of zero where Title V was NA as it was likely NA due to being zero
4. Verified Cost to Charge Ratio calculation as a derived feature. Omitted inconsistent values
5. Derived Medicare and Medicaid ratios based on Total Days. When choosing between correlated substitutes, Total Days was chosen in lieu of Total Discharges as it includes both the inpatient and outpatient visits whereas discharges includes only inpatient visits. These derived features reduce multicollinearity, and help to analyze the role of Medicare and Medicaid health plans
    - Total.Days.XIX.medicaid.ratio = Total Days Title XIX/(Total Days Title V + XVIII +XIX + Unknown)
    - Total.Days.XVIII.medicare.ratio = Total Days Title XVIII/(Total Days Title V + XVIII +XIX + Unknown)
    - Total.Days.unknown.ratio = Total Days Unknown/(Total Days Title V + XVIII +XIX + Unknown)
6. Derived financial leverage ratio to determine the financial risk of a hospital by comparing debt obligations to its assets. This also helps us overcome multicollinearity between financial variables
    - debt.to.asset.ratio = (Total Current + Long Term Liabilities)/Total Assets
7. Removed rows with negative values from Total Income, Debt to Asset Ratio, Total Current and Long-Term Liabilities and Total Unreimbursed and Uncompensated Care
8. Created a data frame subset based on the initial predictor table. Final predictor variables for the models are chosen after examining each highly correlated variable pair in the correlation matrix and making appropriate exclusions based on logical substitutions and derived feature combinations with the objective of avoiding model coefficient bias due to multicollinearity

The resultant working dataset consists of 9,724 observations. Final working predictor variable count is 11.

Explanation of Key Variables

Cost to Charge Ratio is Total Cost/(Inpatient + Outpatient Charges). It does not include revenue from other sources besides that from core patient services.

TITLE V, XVIII & XIX: The dataset includes 35 variables that reference one or more of three Title Types. The Three Title Types are TITLE V, XVIII & XIX. Each of these refer to a different patient class as defined by provisions established in the Social Security Act of 1965.

Title V is a patient classification for whom services provided fall under the provisions of the SSA focused on improving health of mothers and children. Title V is administered by the Maternal and Child Health Bureau within the US Department of Health and Human Services.

Title XVIII is the patient classification for seniors above 65 and younger people with disabilities using federally provided health insurance. It is **commonly known as the Medicare Title**.

Title XIX is the patient classification for those who qualify for federal insurance due to low income and/or certain disability types. **It is commonly known as the Medicaid Title.**

Type of Control indicates the form in which the provider (hospital) has been incorporated. There are 13 unique types of control that indicate various private, non-profit and government types of control.

Provider Type is the specialty care type of the hospital, with five unique values: General Short Term, General Long Term, Rehabilitation, Psychiatric, Cancer.

Descriptive Analytics & Visualizations

Target Variable Distribution:

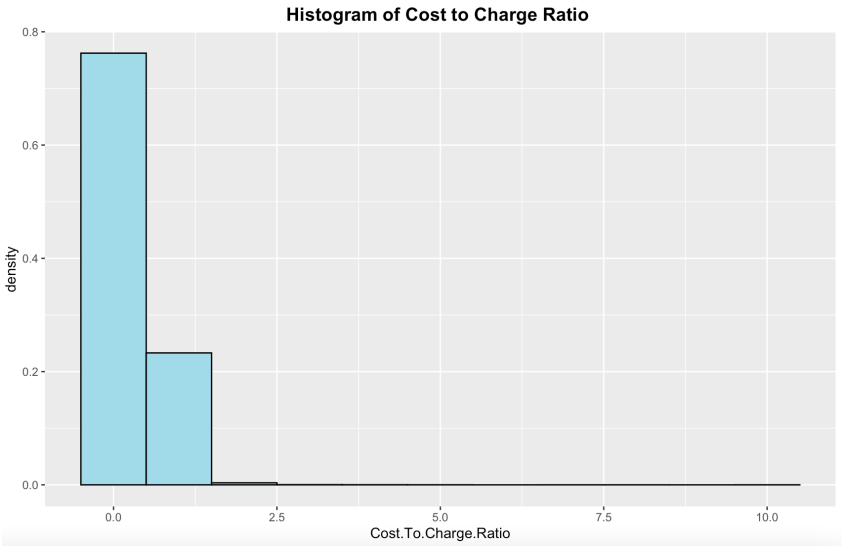


Figure 1 - Histogram: Cost to Charge Ratio

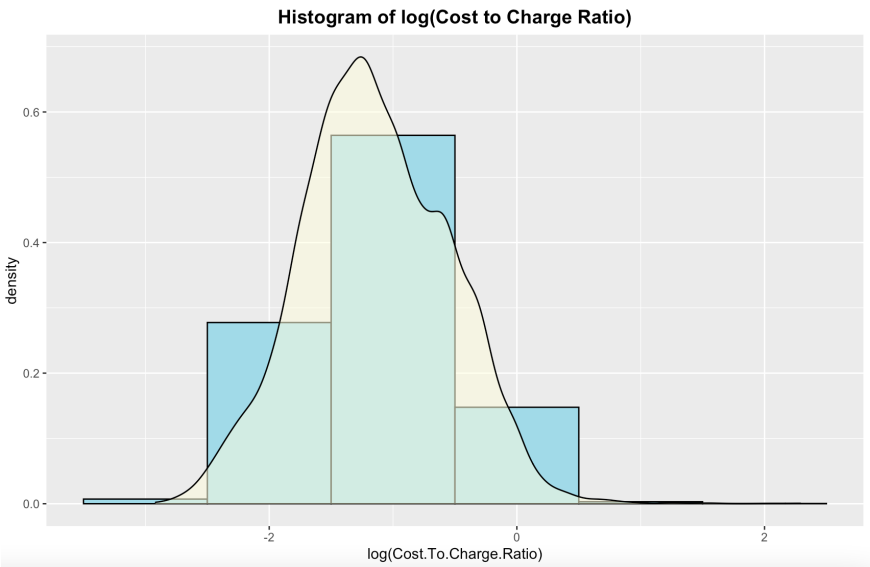


Figure 2 - Histogram log(Cost to Charge Ratio)

We can see that the target variable is not normally distributed, log transform produces a more normal distribution of the target variable. We will use the log transform of the target for analysis to better align with critical model assumptions.

An initial look at a few important factor variables and their relationship with the target

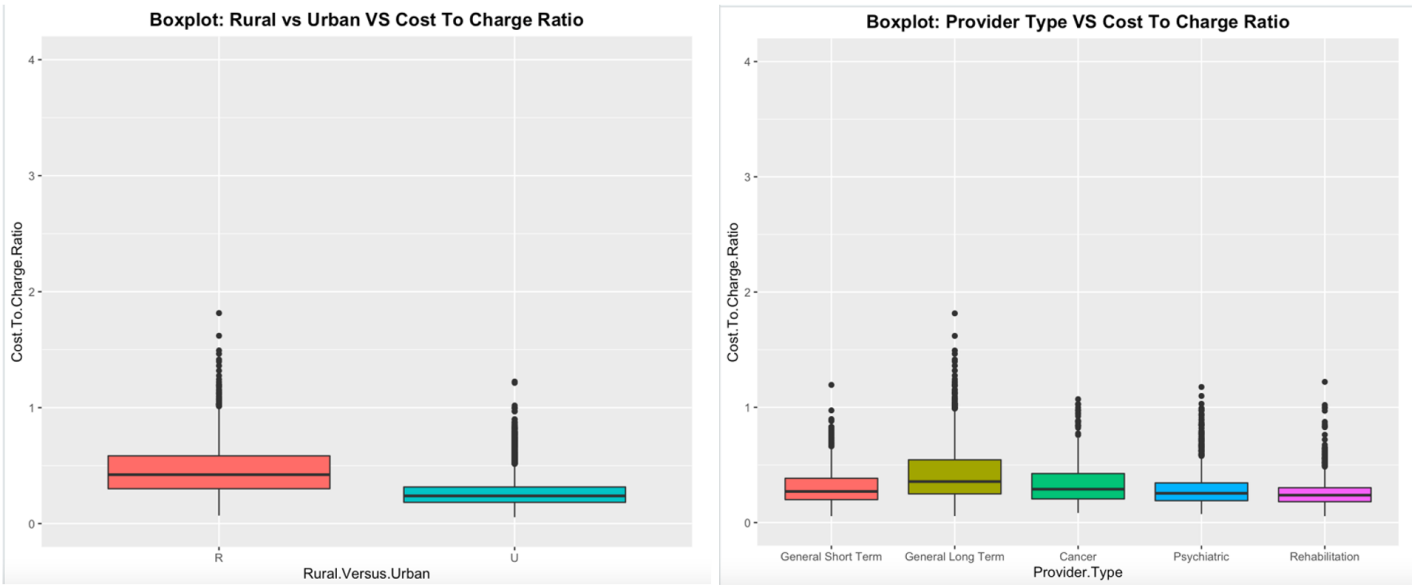


Figure 3 - Boxplots: RvU & Provider Type VS Cost to Charge Ratio

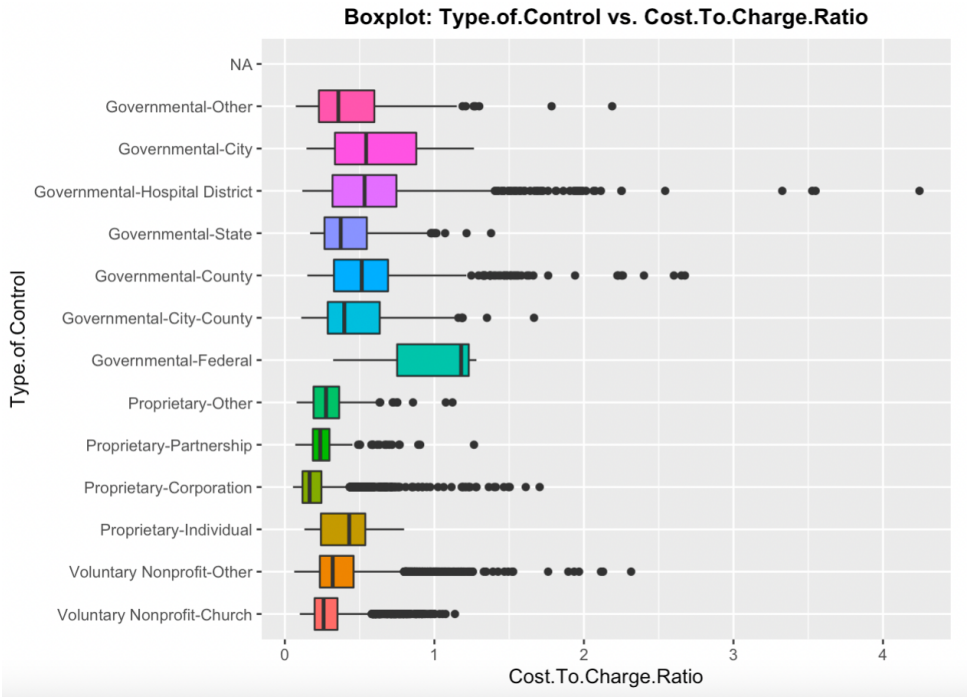


Figure 4 - Boxplot: Type of Control VS Cost to Charge Ratio

We can see clear effects of Rural vs Urban designation, Provider Type, and Type of Control on Cost to Charge Ratio. Government types of control seem to have higher Cost to Charge Ratios while Private control types have lower. These relationships will be explored further and quantified.

Correlation Matrix, Variable exclusion & Initial Predictor Table

Initially we identified 30 variables of interest from the total set of 129 predictors. To avoid multicollinearity, we created a correlation matrix of the numeric variables and identified highly correlated predictor pairs using a threshold of 0.7 for correlation. The final correlation matrix is shown below. The final correlation matrix includes seven numeric variables which form the basis for the subsequent modeling exercise.

	Total.Days.XVIII.medicare.ratio	Total.Days.XIX.medicaid.ratio	Total.Days.unknown.ratio	Total.Unreimbursed.and.Uncompensated.Care	debt.to.asset.ratio	Total.Income	Health.Information.Technology.Designated.Assets
Total.Days.XVIII.medicare.ratio	1	-0.31	-0.84	-0.66	-0.34	-0.12	0.32
Total.Days.XIX.medicaid.ratio	-0.31	1	0.27	0.9	0.61	0.93	0.79
Total.Days.unknown.ratio	-0.84	0.27	1	0.47	-0.14	0.28	-0.33
Total.Unreimbursed.and.Uncompensated.Care	-0.66	0.9	0.47	1	0.75	0.71	0.49
debt.to.asset.ratio	-0.34	0.61	-0.14	0.75	1	0.3	0.49
Total.Income	-0.12	0.93	0.28	0.71	0.3	1	0.8
Health.Information.Technology.Designated.Assets	0.32	0.79	-0.33	0.49	0.49	0.8	1

Table 1 - Final Correlation Matrix

Predictor Table

Variable Name	Effect Cost to Charge Ratio	Rationale
rpt_rec_num	none	The report number should not have any effect on cost to charge ratio
Hospital Name	none	Hospital Name should not have any effect on cost to charge ratio
State Code	+/-	Cost of services & Charges levied will likely vary by state
Rural versus Urban	+	Rural/Urban desgination would affect cost to charge ratio
Provider Type	+/-	Types of service would have an effect on cost to charge ratio as certain services would cost more/bring in more revenue than others
Type of Control	+/-	ToC would affect cost to charge ratio as certain hospitals would be operating for profit and others would not. Also, government hospitals likely have differing reimbursement policies than private hospitals, and could differ by state
FTE - Employees on Payroll	+/-	More FTE employees increases cost, but may also facilitate higher revenue generation
Number of Beds	+/-	More number of beds increases cost, but may also facilitate higher revenue generation
Total Days filed under Title XIX (Medicaid)	+/-	Number of Days can be used as a proxy for number of paitents treated. Increased number of patients would increase both cost and revenue, but maybe in differing amounts
Total Days filed under Title XVIII (Medicare)		
Total Days filed under Title V (Mothers & Children)		
Total Days not filed under SSA (Unknown)		
Cost of charity care	+	Increase in cost of charity care will likely increase cost to charge ratio
Total Bad Debt expense	+	Increase in Bad Debt Expense will likely increases cost to charge ratio
Total Unreimbursed and Uncompensated Care	+	Increase in cost unreimbursed and uncompensated care would increase cost to charge ratio
Overhead Non-Salary Costs	+	Increase in overhead costs would increase cost to charge ratio
Depreciation Cost	+	Increase in depreciation cost will increase cost to charge ratio
Inpatient Total Charges	-	Increase in Inpatient total charges will decrease cost to charge ratio
Outpatient Total Charges	-	Increase in Outpatient total charges will decrease cost to charge ratio
Cash on hand and in banks	+/-	Amount of Cash on Hand may affect how the hospital sets charge rates and how much they prioritize spending which would have an effect on cost to charge ratio
Total Current Assets	+/-	Increase in Hospital's total fixed/current assets would affect cost to charge ratio. ex. having more fixed assets may facilitate hospitals to get loans, which would affect cost to charge ratio
Total fixed Assets		
Health Information Technology Designated Assets	+/-	Increasing value of Hospital IT Assets would correspond to an increased cost, but may also facilitate revenue generation
Total current liabilities	+	Increase in Current Liabilities could increase cost to charge ratio
Total long term liabilities	+	Long term Liabilities could increase cost to charge ratio
Total Income	-	Increased total income would decrease cost to charge ratio
Net Revenue from Medicaid	-	Increased Total Net Revenue from Medicaid would decrease cost to charge ratio

Table 2 - Predictor Table

Statistical Models for Analysis

This section discusses the models built for analysis. Please refer to the appendix to see individual model summary outputs and stargazer model comparison output.

Model 1 - Base Model

The base model was built to evaluate important factor variables' effect on Cost to Charge Ratio. Log transform was applied to the target variable (Cost to Charge Ratio) to induce normality. All interpretations will reference percentage change in Cost to Charge Ratio as a result of the predictors considered.

The base model is a mixed effect model. Type of Control and Provider type are fixed effects, while years and states are random effects in a fixed slope, random intercept model. We chose this model as we would expect years and states to share similar trends (slopes) within their respective groups, but from different "starting" points (intercepts).

```
base_model <- log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban + Provider.Type +
Type.of.Control + (1 | year) + (1 | State.Code) Data: df
```

Model 2 - Medicare and Medicaid

To evaluate the effect of Medicare and Medicaid on Cost to Charge Ratio we add the following derived variables from feature engineering to the base model: Medicaid, Medicare and Unknown Ratios, Debt to Asset Ratio

```
medicare_medicaid_model<-log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban +
Provider.Type +  Type.of.Control + log(Total.Days.XIX.medicaid.ratio)
log(Total.Days.XVIII.medicare.ratio) + log(Total.Days.unknown.ratio) +
log(debt.to.asset.ratio) + (1 | year) + (1 | State.Code)Data: df
```

Model 3 - Health IT Assets

This model analyzes effects of predictors on Cost to Charge Ratio for the subset of hospitals which have invested in Health IT Assets.

```
hit_model<- log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban + Provider.Type +
Type.of.Control + log(Total.Days.XIX.medicaid.ratio) +
log(Total.Days.XVIII.medicare.ratio) + log(Total.Days.unknown.ratio) +
log(Health.Information.Technology.Designated.Assets) +
(debt.to.asset.ratio) + (1 | year) + (1 | State.Code) Data: hit_df
```

Choice of Model

From the 3 models above, **Model 2** is the comprehensive nested model with a focus on key significant factors except Health IT. The impact of Health IT assets is intriguing. However, due to numerous missing values for the feature, we subset the data and built another model to study its effects individually.

Interpretation of marginal effects

Note: we have used color coding in marginal effects tables to indicate direction of effect. Green indicates top three favorable (negative) effects and red indicates top three unfavorable (positive) effects. Yellow denotes effects of key variables of interest.

Model 2 - medicare_medicaid_model

1. **Rural vs Urban:** Urban hospitals have 20.92% lower Cost to Charge Ratio than Rural hospitals.
2. **Type of Control** (with respect to Proprietary- Individual Control): Proprietary Corporations type of control has a negative effect of 25.23% on Cost to Charge Ratio. Hospitals with Governmental Hospital-District, Governmental State, and Governmental County types of control have positive effects of 41.09%, 36.37% and 27.01% respectively on Cost to Charge Ratio.

3. **Medicare, Medicaid, Non-Public:** As Medicare ratio increases by 100% Cost to Charge Ratio decreases by 15.68%. A 100% increase in Medicaid ratio decreases Cost to Charge Ratio by 7.5%. As Unknown (Non Public Insurance/No Insurance) ratio increases by 100%, Cost to Charge Ratio decreases by 33.5%
4. **Random Effects:** State Code and Year have random effects with variances of 0.085 and 0.0007 respectively.

Variable	β	Interpretation value (%)
log(Total.Days.unknown.ratio)	-0.335	-33.496
Type.of.ControlProprietary-Corporation	-0.291	-25.232
Rural.Versus.UrbanU	-0.235	-20.918
log(Total.Days.XVIII.medicare.ratio)	-0.157	-15.678
log(Total.Days.XIX.medicaid.ratio)	-0.076	-7.573
Type.of.ControlGovernmental-County	0.239	27.013
Type.of.ControlGovernmental-State	0.310	36.373
Type.of.ControlGovernmental-HospitalDistrict	0.344	41.088

Table 3 - medicare_medicaid_model Predictor

Model 3 - hit_model

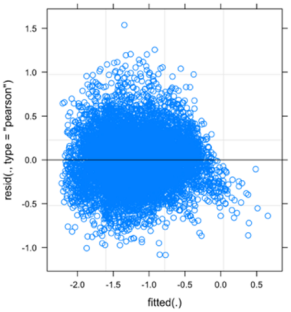
1. **Rural vs Urban:** Urban hospitals have 21.70% lower Cost to Charge Ratio than Rural hospitals.
2. **Type of Control** (with respect to Proprietary - Individual Control): Proprietary Corporations type of control has a negative effect of 48.32% on Cost to Charge Ratio. Governmental State type of control has a positive effect of 19.87% on Cost to Charge Ratio.
3. **Provider Type** (with respect to Provider Type - General Short Term):
   Provider types (Specialty Care type) of Cancer and General Long Term have positive effects of 15.81% and 16.58% on Cost to Charge Ratio respectively
4. **Medicare, Medicaid, Non-Public**: As Medicare ratio increases by 100% Cost to Charge Ratio decreases by 13.14% whereas a 100% increase in Medicaid ratio decreases Cost to Charge Ratio by 5.72%. As Unknown (Non-Public Insurance/No Insurance) ratio increases by 100%, Cost to Charge Ratio decreases by 23.8%
5. **Health Information Technology Assets:** As Health IT Asset value increases by 100%, Cost to Charge Ratio decreases by 2.57%
6. **Random Effects:** State Code and Year have random effects with a variances of 0.079 and 0.0006 respectively.

Variable	β	Interpretation value (%)
Type.of.ControlProprietary-Corporation	-0.660	-48.315
log(Total.Days.unknown.ratio)	-0.238	-23.808
Rural.Versus.UrbanU	-0.245	-21.708
log(Total.Days.XVIII.medicare.ratio)	-0.131	-13.140
log(Total.Days.XIX.medicaid.ratio)	-0.057	-5.719
log(Health.Information.Technology.Designated.Assets)	-0.026	-2.567
Provider.TypeCancer	0.147	15.815
Provider.TypeGeneralLong Term	0.153	16.577
Type.of.ControlGovernmental-State	0.181	19.871

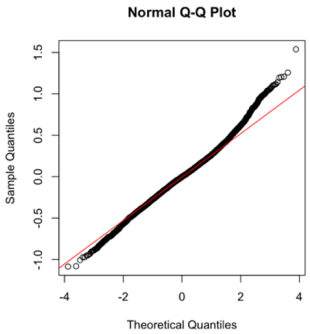
Table 4 - IT Model Predictor Effects

Quality Checks

Model 2: medicare_medicaid_model



Residual plot shows slight downward linear bias, but not enough to consider a failure of assumption



Normality Test: Failed

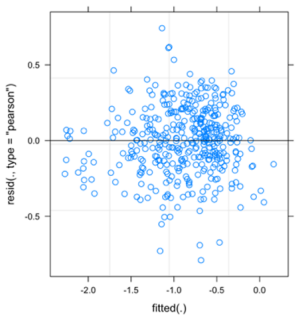
Bartlett's test for Homoskedasticity: Failed
Bartlett's K-squared = 1610.7, df = 1, p-value < 2.2e-16

VIF Test for Multicollinearity: Passed

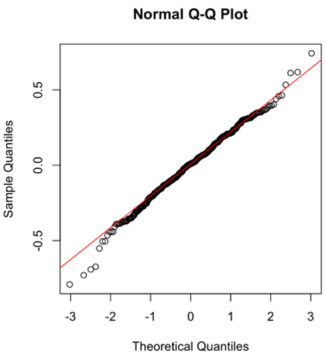
	GVIF	Df	GVIF^(1/(2*Df))
Rural.Versus.Urban	1.358699	1	1.165633
Provider.Type	1.113166	4	1.013491
Type.of.Control	1.258507	11	1.010506
log(Total.Days.XIX.m	1.22098	1	1.104979
log(Total.Days.XVIII.ı	1.865235	1	1.365736
log(Total.Days.unknc	1.837397	1	1.355506
log(debt.to.asset.rat	1.03636	1	1.018018

Figure 5 - Model 2 Quality Check

Model 3: hit_model



Residual plot shows no major issues with linearity



Normality Test: Failed

Bartlett's test for Homoskedasticity: Failed
Bartlett's K-squared = 176.03, df = 1, p-value < 2.2e-16

VIF Test for Multicollinearity: Passed

	GVIF	Df GVI	F^(1/(2*Df))
Rural.Versus.Urban	1.406787	1	1.18608
Provider.Type	1.347536	4	1.037988
Type.of.Control	1.716267	10	1.027376
log(Total.Days.XIX.m	1.285569	1	1.13383
log(Total.Days.XVIII.ı	2.218901	1	1.489598
log(Total.Days.unknc	2.18567	1	1.478401
log(Health.Informati	1.276489	1	1.129818
log(debt.to.asset.rat	1.124291	1	1.060326

Figure 6 - Model 3 Quality Check

Recommendations:

Based on the analysis of Models 2 & 3 we present the following recommendations.

- Type of Control was the most impactful predictor of Cost to Charge Ratio. Private hospitals had a much more favorable effect than government hospitals. Therefore, we recommend the following
  1. Government hospitals should adopt applicable private sector operational strategies.
  2. Invest in Health IT to streamline adopted strategies - Coefficient comparison between models 2 and 3 shows that hospitals that invested in Health IT had significant percentage point reductions in their effect on Cost to Charge Ratio. The effect of County controlled hospitals changed direction from +27.01% to -15.92%

- When compared to rural hospitals, urban hospitals have a 20.91% lower Cost to Charge Ratio. In the short term, to address the immediate threat of rural hospital closings, we recommend prioritizing reimbursements to rural hospitals.

- Our analysis found that as both Medicare and Medicaid ratios increase, Cost to Charge Ratio decreases. However, the reduction is much greater for Medicare vs Medicaid ratio (15.68% vs. 7.57%)

  Medicare is known to have both better coverage and reimbursement policies than Medicaid. Both of these advantages likely contribute to the greater favorable effect on Cost-to-Charge ratio. Since broadening coverage for Medicaid may be fiscally infeasible in the short term, Medicaid should at least adopt reimbursement policies like those of Medicare.

References

- Bai, G., & Anderson, G. F. (2015). Extreme markup: The Fifty US hospitals with the highest charge-to-cost ratios. *Health Affairs*, *34*(6), 922–928. https://doi.org/10.1377/hlthaff.2014.1414
- Bai, G., & Anderson, G. F. (2016). A more detailed understanding of factors associated with hospital profitability. *Health Affairs*, *35*(5), 889–897. https://doi.org/10.1377/hlthaff.2015.1193
- Balasubramanian, S., & Jones, E. (2016). Hospital closures and the current healthcare climate: The future of rural hospitals in the USA. *Rural and Remote Health*. https://doi.org/10.22605/rrh3935
- Carey, K. (2003). Hospital cost efficiency and system membership. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *40*(1), 25–38. https://doi.org/10.5034/inquiryjrnl_40.1.25
- Ferraris, V. A., Ferraris, S. P., & Singh, A. (1998). Operative outcome and hospital cost. *The Journal of Thoracic and Cardiovascular Surgery*, *115*(3), 593–603. https://doi.org/10.1016/s0022-5223(98)70324-1
- Hassanain, M. A., Assaf, S., Al-Ofi, K., & Al-Abdullah, A. (2013). Factors affecting maintenance cost of hospital facilities in Saudi Arabia. *Property Management*, *31*(4), 297–310. https://doi.org/10.1108/pm-10-2012-0035
- Thornton, J. A., & Beilfuss, S. N. (2015). New evidence on factors affecting the level and growth of US Health Care Spending. *Applied Economics Letters*, *23*(1), 15–18. https://doi.org/10.1080/13504851.2015.1044644
- Thornton, J. A., & Rice, J. L. (2008). Determinants of healthcare spending: A State level analysis. *Applied Economics*, *40*(22), 2873–2889. https://doi.org/10.1080/00036840600993973

Appendix

Parting Thoughts:

This analysis revealed many important lessons in statistical modeling for the purpose of providing actionable insights. We found that even federally compiled, official reporting data is wrought with inconsistencies, missing values and infeasible figures. Working effectively with raw datasets requires a deep analysis of all variables before model building. Data cleaning, feature engineering and selection are arguably the most critical components of the process. Without an effective data cleanse, modeling output could produce inaccurate insights. We strongly believe that aspiring Data Science professionals should be aware of the negative consequences of acting upon inaccurate insights, and thus should proceed with an attitude of thoroughness and attention to detail.

Correlation Matrices

	FTE...Employees.on.Payroll	Number.of.Beds	Total.Days.XVIII.medicare.ratio	Total.Days.XIX.medicaid.ratio	Total.Days.V.ratio	Total.Days.unknown.ratio	Total.Bed.Days.Available	Total.Discharges.Title.V	Total.Discharges.Title.XVIII	Total.Discharges.Title.XIX	Cost.of.Charity.Care	Total.Bad.Debt.Expense	Total.Unreimbursed.and.Uncompensated.Care	Overhead.Non.Salary.Costs	Depreciation.Cost	Inpatient.Total.Charges	Outpatient.Total.Charges	Total.Salaries..adjusted.	Cash.on.Hand.and.in.Banks	Total.Current.Assets	Total.fixed.Assets	Total.Assets	Inpatient.Revenue	Outpatient.Revenue	Less.Total.Operating.Expense	Total.Income	Cost.To.Charge.Ratio	debt.to.asset.ratio	Net.Revenue.from.Medicaid	Health.Information.Technology.Designated.Assets
FTE...Employees.on.Payroll	1	0.81	-0.36	1	-0.48	0.29	0.8	-0.51	0.52	0.62	0.92	0.76	0.92	1	0.98	0.91	1	0.98	0.58	0.91	0.73	1	0.91	1	0.98	0.91	0.92	0.65	0.94	0.76
Number.of.Beds	0.81	1	-0.52	0.81	-0.9	0.72	1	-0.81	0.45	0.55	0.73	0.31	0.75	0.82	0.89	0.74	0.86	0.92	0.76	0.76	0.68	0.82	0.72	0.85	0.81	0.87	0.88	0.14	0.74	0.41
Total.Days.XVIII.medicare.ratio	-0.36	-0.52	1	-0.31	0.34	-0.84	-0.51	0.89	0.52	0.4	-0.67	-0.44	-0.66	-0.42	-0.47	0.02	-0.36	-0.44	-0.95	-0.7	0.24	-0.32	0.03	-0.34	-0.21	-0.12	-0.17	-0.34	-0.62	0.32
Total.Days.XIX.medicaid.ratio	1	0.81	-0.31	1	-0.51	0.27	0.81	-0.48	0.57	0.67	0.89	0.72	0.9	0.99	0.98	0.94	1	0.98	0.54	0.88	0.77	1	0.94	1	0.99	0.93	0.95	0.61	0.91	0.79
Total.Days.V.ratio	-0.48	-0.9	0.34	-0.51	1	-0.75	-0.9	0.73	-0.46	-0.52	-0.37	0.14	-0.39	-0.5	-0.61	-0.52	-0.57	-0.66	-0.59	-0.4	-0.62	-0.53	-0.5	-0.56	-0.53	-0.71	-0.7	0.31	-0.36	-0.18
Total.Days.unknown.ratio	0.29	0.72	-0.84	0.27	-0.75	1	0.72	-0.97	-0.22	-0.13	0.46	0.02	0.47	0.35	0.45	0.07	0.35	0.47	0.89	0.51	0.02	0.29	0.05	0.32	0.22	0.28	0.31	-0.14	0.41	-0.33
Total.Bed.Days.Available	0.8	1	-0.51	0.81	-0.9	0.72	1	-0.81	0.45	0.56	0.73	0.31	0.75	0.82	0.89	0.74	0.85	0.92	0.76	0.76	0.68	0.82	0.72	0.85	0.81	0.87	0.88	0.14	0.73	0.41
Total.Discharges.Title.V	-0.51	-0.81	0.89	-0.48	0.73	-0.97	-0.81	1	0.15	0.03	-0.67	-0.26	-0.68	-0.56	-0.65	-0.25	-0.55	-0.65	-0.97	-0.71	-0.13	-0.5	-0.23	-0.52	-0.42	-0.43	-0.47	-0.1	-0.63	0.14
Total.Discharges.Title.XVIII	0.52	0.45	0.52	0.57	-0.46	-0.22	0.45	0.15	1	0.99	0.13	0	0.15	0.47	0.46	0.82	0.55	0.51	-0.22	0.11	0.95	0.57	0.82	0.57	0.66	0.78	0.75	-0.05	0.19	0.84
Total.Discharges.Title.XIX	0.62	0.55	0.4	0.67	-0.52	-0.13	0.56	0.03	0.99	1	0.26	0.09	0.28	0.58	0.57	0.89	0.65	0.62	-0.09	0.24	0.99	0.67	0.89	0.67	0.75	0.86	0.83	0.03	0.31	0.87
Cost.of.Charity.Care	0.92	0.73	-0.67	0.89	-0.37	0.46	0.73	-0.67	0.13	0.26	1	0.87	1	0.94	0.93	0.67	0.9	0.89	0.79	1	0.4	0.89	0.67	0.89	0.83	0.69	0.73	0.76	1	0.48
Total.Bad.Debt.Expense	0.76	0.31	-0.44	0.72	0.14	0.02	0.31	-0.26	0	0.09	0.87	1	0.85	0.76	0.69	0.51	0.69	0.63	0.46	0.84	0.18	0.71	0.52	0.69	0.66	0.42	0.46	0.98	0.87	0.51
Total.Unreimbursed.and.Uncompensated.Care	0.92	0.75	-0.66	0.9	-0.39	0.47	0.75	-0.68	0.15	0.28	1	0.85	1	0.94	0.94	0.69	0.91	0.91	0.79	1	0.42	0.9	0.68	0.9	0.84	0.71	0.74	0.75	1	0.49
Overhead.Non.Salary.Costs	1	0.82	-0.42	0.99	-0.5	0.35	0.82	-0.56	0.47	0.58	0.94	0.76	0.94	1	0.99	0.89	0.99	0.98	0.63	0.93	0.69	0.99	0.89	0.99	0.97	0.89	0.91	0.65	0.95	0.72
Depreciation.Cost	0.98	0.89	-0.47	0.98	-0.61	0.45	0.89	-0.65	0.46	0.57	0.93	0.69	0.94	0.99	1	0.87	0.99	1	0.7	0.93	0.7	0.98	0.87	0.99	0.96	0.91	0.93	0.56	0.94	0.66
Inpatient.Total.Charges	0.91	0.74	0.02	0.94	-0.52	0.07	0.74	-0.25	0.82	0.89	0.67	0.51	0.69	0.89	0.87	1	0.92	0.89	0.27	0.66	0.94	0.94	1	0.93	0.97	0.97	0.97	0.42	0.72	0.92
Outpatient.Total.Charges	1	0.86	-0.36	1	-0.57	1	0.35	0.85	-0.55	0.55	0.65	0.9	0.69	0.91	0.99	0.99	0.92	1	0.99	0.6	0.89	0.76	1	0.92	1	0.99	0.94	0.95	0.57	0.92	0.75
Total.Salaries..adjusted.	0.98	0.92	-0.44	0.98	-0.66	0.47	0.92	-0.65	0.51	0.62	0.89	0.63	0.91	0.98	1	0.89	0.99	1	0.68	0.9	0.74	0.98	0.88	0.99	0.96	0.93	0.95	0.49	0.91	0.66
Cash.on.Hand.and.in.Banks	0.58	0.76	-0.95	0.54	-0.59	0.89	0.76	-0.97	-0.22	-0.09	0.79	0.46	0.79	0.63	0.7	0.27	0.6	0.68	1	0.82	0.07	0.56	0.25	0.58	0.47	0.42	0.46	0.32	0.75	-0.08
Total.Current.Assets	0.91	0.76	-0.7	0.88	-0.4	0.51	0.76	-0.71	0.11	0.24	1	0.84	1	0.93	0.93	0.66	0.89	0.9	0.82	1	0.39	0.88	0.65	0.88	0.82	0.69	0.73	0.73	0.99	0.44
Total.fixed.Assets	0.73	0.68	0.24	0.77	-0.62	0.02	0.68	-0.13	0.95	0.99	0.4	0.18	0.42	0.69	0.7	0.94	0.76	0.74	0.07	0.39	1	0.77	0.93	0.77	0.84	0.93	0.91	0.09	0.45	0.87
Total.Assets	1	0.82	-0.32	1	-0.53	0.29	0.82	-0.5	0.57	0.67	0.89	0.71	0.9	0.99	0.98	0.94	1	0.98	0.56	0.88	0.77	1	0.93	1	0.99	0.93	0.95	0.6	0.91	0.78
Inpatient.Revenue	0.91	0.72	0.03	0.94	-0.5	0.05	0.72	-0.23	0.82	0.89	0.67	0.52	0.68	0.89	0.87	1	0.92	0.88	0.25	0.65	0.93	0.93	1	0.93	0.97	0.97	0.96	0.43	0.71	0.93
Outpatient.Revenue	1	0.85	-0.34	1	-0.56	0.32	0.85	-0.52	0.57	0.67	0.89	0.69	0.9	0.99	0.99	0.93	1	0.99	0.58	0.88	0.77	1	0.93	1	0.99	0.94	0.96	0.57	0.91	0.77
Less.Total.Operating.Expense	0.98	0.81	-0.21	0.99	-0.53	0.22	0.81	-0.42	0.66	0.75	0.83	0.66	0.84	0.97	0.96	0.97	0.99	0.96	0.47	0.82	0.84	0.99	0.97	0.99	1	0.96	0.97	0.55	0.86	0.84
Total.Income	0.91	0.87	-0.12	0.93	-0.71	0.28	0.87	-0.43	0.78	0.86	0.69	0.42	0.71	0.89	0.91	0.97	0.94	0.93	0.42	0.69	0.93	0.93	0.97	0.94	0.96	1	1	0.3	0.72	0.8
Cost.To.Charge.Ratio	0.92	0.88	-0.17	0.95	-0.7	0.31	0.88	-0.47	0.75	0.83	0.73	0.46	0.74	0.91	0.93	0.97	0.95	0.95	0.46	0.73	0.91	0.95	0.96	0.96	0.97	1	1	0.33	0.76	0.79
debt.to.asset.ratio	0.65	0.14	-0.34	0.61	0.31	-0.14	0.14	-0.1	-0.05	0.03	0.76	0.98	0.75	0.65	0.56	0.42	0.57	0.49	0.32	0.73	0.09	0.6	0.43	0.57	0.55	0.3	0.33	1	0.77	0.49
Net.Revenue.from.Medicaid	0.94	0.74	-0.62	0.91	-0.36	0.41	0.73	-0.63	0.19	0.31	1	0.87	1	0.95	0.94	0.72	0.92	0.91	0.75	0.99	0.45	0.91	0.71	0.91	0.86	0.72	0.76	0.77	1	0.54
Health.Information.Technology.Designated.Assets	0.76	0.41	0.32	0.79	-0.18	-0.33	0.41	0.14	0.84	0.87	0.48	0.51	0.49	0.72	0.66	0.92	0.75	0.66	-0.08	0.44	0.87	0.78	0.93	0.77	0.84	0.8	0.79	0.49	0.54	1

Table 5 - Initial Correlation Matrix

	Total.Days.XVIII.medicare.ratio	Total.Days.XIX.medicaid.ratio	Total.Days.unknown.ratio	Total.Unreimbursed.and.Uncompensated.Care	debt.to.asset.ratio	Total.Income	Health.Information.Technology.Designated.Assets
Total.Days.XVIII.medicare.ratio	1	-0.31	-0.84	-0.66	-0.34	-0.12	0.32
Total.Days.XIX.medicaid.ratio	-0.31	1	0.27	0.9	0.61	0.93	0.79
Total.Days.unknown.ratio	-0.84	0.27	1	0.47	-0.14	0.28	-0.33
Total.Unreimbursed.and.Uncompensated.Care	-0.66	0.9	0.47	1	0.75	0.71	0.49
debt.to.asset.ratio	-0.34	0.61	-0.14	0.75	1	0.3	0.49
Total.Income	-0.12	0.93	0.28	0.71	0.3	1	0.8
Health.Information.Technology.Designated.Assets	0.32	0.79	-0.33	0.49	0.49	0.8	1

Table 6 - Final Correlation Matrix

Model Summaries

Base_model

```
> summary(base_model)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban + Provider.Type +
Type.of.Control + (1 | year) + (1 | State.Code)
   Data: df

     AIC      BIC   logLik deviance df.resid
  5187.9   5331.6  -2574.0   5147.9     9704

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8365 -0.6329 -0.0089  0.5872  6.1882

Random effects:
 Groups     Name        Variance  Std.Dev.
 State.Code (Intercept) 0.0889316 0.29821
 year       (Intercept) 0.0005447 0.02334
 Residual               0.0967561 0.31106
Number of obs: 9724, groups:  State.Code, 53; year, 5

Fixed effects:
                                              Estimate Std. Error t value
(Intercept)                                  -0.961571   0.075601 -12.719
Rural.Versus.UrbanU                          -0.305521   0.007577 -40.324
Provider.TypeGeneral Long Term                0.113033   0.010360  10.910
Provider.TypeCancer                           0.047936   0.016169   2.965
Provider.TypePsychiatric                      0.031743   0.013735   2.311
Provider.TypeRehabilitation                  -0.044762   0.011631  -3.848
Type.of.ControlVoluntary Nonprofit-Church    -0.127918   0.062477  -2.047
Type.of.ControlVoluntary Nonprofit-Other     -0.024319   0.061968  -0.392
Type.of.ControlProprietary-Corporation       -0.352935   0.062490  -5.648
Type.of.ControlProprietary-Partnership       -0.110205   0.067384  -1.635
Type.of.ControlProprietary-Other             -0.163676   0.068820  -2.378
Type.of.ControlGovernmental-City-County       0.147003   0.065927   2.230
Type.of.ControlGovernmental-County            0.218662   0.063055   3.468
Type.of.ControlGovernmental-State             0.227036   0.069368   3.273
Type.of.ControlGovernmental-Hospital District 0.314733   0.063158   4.983
Type.of.ControlGovernmental-City              0.133808   0.068610   1.950
Type.of.ControlGovernmental-Other             0.103080   0.068086   1.514

> ## multicollinearity test
> vif(base_model)
                       GVIF Df GVIF^(1/(2*Df))
Rural.Versus.Urban 1.152979  1        1.073768
Provider.Type      1.080617  4        1.009739
Type.of.Control    1.134946 11        1.005770
```

Medicare_medicaid_model

```
> summary(medicare_medicaid_model)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban + Provider.Type +
    Type.of.Control + log(Total.Days.XIX.medicaid.ratio) +
log(Total.Days.XVIII.medicare.ratio) +
    log(Total.Days.unknown.ratio) + log(debt.to.asset.ratio) + (1 | year) + (1 |
State.Code)
   Data: df

     AIC      BIC   logLik deviance df.resid
  3715.3   3887.7  -1833.7   3667.3     9700

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7679 -0.6303 -0.0040  0.5978  5.3373

Random effects:
 Groups     Name        Variance  Std.Dev.
 State.Code (Intercept) 0.0852471 0.29197
 year       (Intercept) 0.0007659 0.02767
 Residual               0.0830255 0.28814
Number of obs: 9724, groups:  State.Code, 53; year, 5

Fixed effects:
                                             Estimate Std. Error t value
(Intercept)                                -1.669365   0.074692 -22.350
Rural.Versus.UrbanU                        -0.234682   0.007620 -30.798
Provider.TypeGeneral Long Term              0.071842   0.009741   7.376
Provider.TypeCancer                         0.037673   0.015051   2.503
Provider.TypePsychiatric                    0.015385   0.012855   1.197
Provider.TypeRehabilitation                -0.031005   0.010793  -2.873
Type.of.ControlVoluntary Nonprofit-Church  -0.033648   0.057939  -0.581
Type.of.ControlVoluntary Nonprofit-Other    0.051367   0.057449   0.894
Type.of.ControlProprietary-Corporation     -0.290775   0.057942  -5.018
Type.of.ControlProprietary-Partnership     -0.083104   0.062578  -1.328
Type.of.ControlProprietary-Other           -0.167915   0.063845  -2.630
Type.of.ControlGovernmental-City-County     0.180722   0.061117   2.957
Type.of.ControlGovernmental-County          0.239117   0.058450   4.091
Type.of.ControlGovernmental-State           0.310220   0.064450   4.813
Type.of.ControlGovernmental-Hospital District  0.344213   0.058541   5.880
Type.of.ControlGovernmental-City            0.214413   0.063628   3.370
Type.of.ControlGovernmental-Other           0.147422   0.063117   2.336
log(Total.Days.XIX.medicaid.ratio)         -0.075732   0.003016 -25.113
log(Total.Days.XVIII.medicare.ratio)       -0.156778   0.008579 -18.275
log(Total.Days.unknown.ratio)              -0.334962   0.010079 -33.235
log(debt.to.asset.ratio)                    0.028755   0.003925   7.326
```

Hit_model

```
> summary(hit_model)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: log(Cost.To.Charge.Ratio) ~ Rural.Versus.Urban + Provider.Type +
    Type.of.Control + log(Total.Days.XIX.medicaid.ratio) +
log(Total.Days.XVIII.medicare.ratio) +
    log(Total.Days.unknown.ratio) +
log(Health.Information.Technology.Designated.Assets) +
    log(debt.to.asset.ratio) + (1 | year) + (1 | State.Code)
   Data: hit_df

     AIC      BIC   logLik deviance df.resid
   137.7    233.4    -44.9     89.7      374

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3466 -0.5656  0.0401  0.6411  3.1414

Random effects:
 Groups     Name        Variance  Std.Dev.
 State.Code (Intercept) 0.0786203 0.28039
 year       (Intercept) 0.0005634 0.02374
 Residual               0.0557827 0.23618
Number of obs: 398, groups:  State.Code, 46; year, 5

Fixed effects:
                                                  Estimate Std. Error t value
(Intercept)                                       -0.87990    0.27387  -3.213
Rural.Versus.UrbanU                               -0.24473    0.03876  -6.315
Provider.TypeGeneral Long Term                     0.15338    0.05510   2.784
Provider.TypeCancer                                0.14682    0.07775   1.888
Provider.TypePsychiatric                           0.03977    0.06345   0.627
Provider.TypeRehabilitation                        0.03808    0.07786   0.489
Type.of.ControlVoluntary Nonprofit-Church         -0.21839    0.18963  -1.152
Type.of.ControlVoluntary Nonprofit-Other          -0.21514    0.17932  -1.200
Type.of.ControlProprietary-Corporation            -0.66001    0.18546  -3.559
Type.of.ControlProprietary-Partnership            -0.21274    0.22333  -0.953
Type.of.ControlGovernmental-City-County           -0.17347    0.19521  -0.889
Type.of.ControlGovernmental-County                -0.11824    0.18575  -0.637
Type.of.ControlGovernmental-State                  0.18125    0.25979   0.698
Type.of.ControlGovernmental-Hospital District      0.12168    0.18609   0.654
Type.of.ControlGovernmental-City                   0.11819    0.20445   0.578
Type.of.ControlGovernmental-Other                  0.01420    0.22595   0.063
log(Total.Days.XIX.medicaid.ratio)                -0.05719    0.01320  -4.331
log(Total.Days.XVIII.medicare.ratio)              -0.13140    0.04911  -2.675
log(Total.Days.unknown.ratio)                     -0.23808    0.04754  -5.008
log(Health.Information.Technology.Designated.Assets) -0.02567    0.01100  -2.333
log(debt.to.asset.ratio)                           0.01705    0.01959   0.870
```

Stargazer Output of Models for Comparison

(1) - base_model, (2) - medicare_medicaid_model , (3) - hit_model

Model Results

	Dependent variable:		
	log(Cost.To.Charge.Ratio)		
	(1)	(2)	(3)
Rural.Versus.UrbanU	-0.306*** (0.008)	-0.235*** (0.008)	-0.245*** (0.039)
Provider.TypeGeneral Long Term	0.113*** (0.010)	0.072*** (0.010)	0.153*** (0.055)
Provider.TypeCancer	0.048*** (0.016)	0.038** (0.015)	0.147* (0.078)
Provider.TypePsychiatric	0.032** (0.014)	0.015 (0.013)	0.040 (0.063)
Provider.TypeRehabilitation	-0.045*** (0.012)	-0.031*** (0.011)	0.038 (0.078)
Type.of.ControlVoluntary Nonprofit-Church	-0.128** (0.062)	-0.034 (0.058)	-0.218 (0.190)
Type.of.ControlVoluntary Nonprofit-Other	-0.024 (0.062)	0.051 (0.057)	-0.215 (0.179)
Type.of.ControlProprietary-Corporation	-0.353*** (0.062)	-0.291*** (0.058)	-0.660*** (0.185)
Type.of.ControlProprietary-Partnership	-0.110 (0.067)	-0.083 (0.063)	-0.213 (0.223)
Type.of.ControlProprietary-Other	-0.164** (0.069)	-0.168*** (0.064)	
Type.of.ControlGovernmental-City-County	0.147** (0.066)	0.181*** (0.061)	-0.173 (0.195)
Type.of.ControlGovernmental-County	0.219*** (0.063)	0.239*** (0.058)	-0.118 (0.186)
Type.of.ControlGovernmental-State	0.227*** (0.069)	0.310*** (0.064)	0.181 (0.260)
Type.of.ControlGovernmental-Hospital District	0.315*** (0.063)	0.344*** (0.059)	0.122 (0.186)
Type.of.ControlGovernmental-City	0.134* (0.069)	0.214*** (0.064)	0.118 (0.204)
Type.of.ControlGovernmental-Other	0.103 (0.068)	0.147** (0.063)	0.014 (0.226)
log(Total.Days.XIX.medicaid.ratio)		-0.076*** (0.003)	-0.057*** (0.013)
log(Total.Days.XVIII.medicare.ratio)		-0.157*** (0.009)	-0.131*** (0.049)
log(Total.Days.unknown.ratio)		-0.335*** (0.010)	-0.238*** (0.048)
log(Health.Information.Technology.Designated.Assets)			-0.026** (0.011)
log(debt.to.asset.ratio)		0.029*** (0.004)	0.017 (0.020)
Constant	-0.962*** (0.076)	-1.669*** (0.075)	-0.880*** (0.274)
Observations	9,724	9,724	398
Log Likelihood	-2,573.951	-1,833.653	-44.851
Akaike Inf. Crit.	5,187.903	3,715.305	137.701
Bayesian Inf. Crit.	5,331.550	3,887.682	233.376
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7 - Stargazer Output of Models

Prior Work & Its Relevance To Our Investigation

Paper	Relevance	Details/key points	Other Notes
Balasubramanian, S., & Jones, E. (2016). **Hospital closures and the current healthcare climate: The future of rural hospitals in the USA**. *Rural and Remote Health*. https://doi.org/10.22605/rrh3935	CONTEXT	1. Hospital Closings are largely due to financial difficulties 2. Medicaid Expansion has increased total number of people covered by 17 million 3. Hospital Readmission Reduction Program (HRRP) and Hospital Value Based Purchasing Program (HVBP) important to improving inpatient care 4. Half the population lives in rural areas 5. 65 Rural hospitals have closed since 2010, 283 more at risk 6. Rural hospital closings have amplified effect as alternate hospitals may be very far away, 7. HRRP and HVBP measures may be hurting hospitals more than improving paitent care, especially in rural areas 8. Medicare accounts for 45% of Rural hospital income - Reimbursement cuts are a problem 9. By law, when a patient enters emergency care, regardless of the patient's ability to pay, the hospital must provide treatment for the patient until they become stabilized or they die. These scenarios incur a debt to hospitals, which are responsible for those costs. 10. events may have expedited the closure of already struggling hospitals with reimbursements cuts and other implications of the law. 11. Hospitals that depend on government reimbursements for the majority of their revenue and perform poorly on inpatient care may be affected.	- Clearly indicates major issues with financial solvency causing rural hospital closings and some of the major factors that are affecting them - May want to explore cost of ED in rural hospitals vs Urban hospitals - Could potentially include data about effectiveness of relevant ACA provisions such as HRRP and HVBP
Bai, G., & Anderson, G. F. (2016). **A more detailed understanding of factors associated with hospital profitability.** *Health Affairs*, *35*(5), 889–897. https://doi.org/10.1377/hlthaff.2015.1193	CONTEXT + METHODOLOGY	1. System affiliation, medicare percentage, prestige seem to be important factors to consider 2. factors largely beyond hospitals' control—location, patient mix, and the relative market power of private insurers—were associated with hospital profitability 3. Factors within hospitals' control, markup and consolidation, also play important roles. 4. "Adjusted discharge" variable is the number of discharges multiplied by the ratio of total gross revenue to inpatient gross revenue and multiplied by the case-mix index and wage index 5. Median Hospital lost $82 per discharge	- Provides strong foundation for which variables we should consider and why - Factors beyond hospital control can possibly be included in our analysis as we have location info of hospitals (create MLM with geography) - May want to explore why/how top 10 hospitals are doing so well

Bai, G., & Anderson, G. F. (2015). E**xtreme markup: The Fifty US hospitals with the highest charge-to-cost ratios.** *Health Affairs*, 34(6), 922–928. https://doi.org/10.1377/hlthaff.2014.1414	CONTEXT	1. Federal and state governments may want to consider limitations on the charge to-cost ratio, some form of all-payer rate setting, or mandated price disclosure to regulate hospital markups 2. hospital markups (ratios of charges over Medicare-allowable costs) do not have an effect on the amounts publicly insured patients pay because Medicare and Medicaid determine their own rules for paying hospital 3.The charge-to-cost ratio varies for inpatient and outpatient care 4.If a hospital offers a discount to certain categories of uninsured patients, its cost report does not report this information 5. privately insured patients may also pay a greater premium because high markups give hospitals greater bargaining power. As a result, high markups play a role in the rise of overall health care spending 6. hospitals need to receive sufficient revenue to remain in business, and having revenues that are above costs is necessary 7. the average charge-to-cost ratio for anesthesiology is 112, for diagnostic radiology it is 15, and for nursery it is 3. To overcome this limitation, one option is to require all hospitals to use a uniform charge-to-cost ratio for all services and disclose this ratio. This approach, by reducing the variation of markups across services, would make it easier for patients to compare hospital prices. 8. Existing laws in some states use a variant of this approach to protect uninsured patients against high hospital charges.	
Carey, K. (2003). **Hospital cost efficiency and system membership.** INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 40(1), 25–38. https://doi.org/10.5034/inquiryjrnl_40.1.25	CONTEXT + METHODOLOGY	1. A number of studies found system-affiliated hospitals to have higher costs per case than freestanding hospitals 2. hospitals belonging to moderately centralized systems performed better than highly centralized systems on measures of hospital costs 3. Unilateral decision making and the development of institutional rules that do not account for local contingencies may have negative production consequences 4. Hospital cost functions are commonly specified using the translog model.	

Thornton, J. A., & Rice, J. L. (2008). **Determinants of healthcare spending: A State level analysis**. Applied Economics, 40(22), 2873–2889. https://doi.org/10.1080/00036840600993973	CONTEXT + METHODOLOGY	1. Methodology is multilevel modeling approach including exogenous variables that affect spending both indirectly, directly and mixed 2. Suggests that mechanism of healthcare spending is product of demand which itself may be driven by ability to spend on healthcare (i.e recursive effect) 3. Explores lifestyle, socio-economic and environmental effects as indirect predictors	May want to consider including lifestyle factors split by geography (i.e. Alcohol Consumption, Cigarette Smoking etc.) - Used Log tranform in OLS Model

Thornton, J. A., & Beilfuss, S. N. (2015). **New evidence on factors affecting the level and growth of US Health Care Spending**. Applied Economics Letters, 23(1), 15–18. https://doi.org/10.1080/13504851.2015.1044644	CONTEXT	1. 27% of Hospital maintenances depends on **Technology Upgradation** 2. The charge of the hospital depends on quality of physicians have the biggest effect (0.51), followed by the elderly (0.27), income (0.24), alcohol consumption (0.23), the black population (0.08), non-HMO health plans (0.06), Medicaid (0.03) and hospitals (0.02) 3. Charges depends on more doctors per capita accounted in the hospital	- Uses log transform OLS model - Raises question of how we can include technology upgrades in cost-charge modeling - Number of doctors per patient may be a good variable to create

Hassanain, M. A., Assaf, S., Al-Ofi, K., & Al-Abdullah, A. (2013). **Factors affecting maintenance cost of hospital facilities in Saudi Arabia.** Property Management, 31(4), 297–310. https://doi.org/10.1108/pm-10-2012-0035	CONTEXT + METHODOLOGY	1.Factors affecting the maintenance of hospital were categorized in seven groups, namely statutory requirements, design phase, construction phase, management of the maintenance department, budgetary estimates for maintenance activities, operations conducted by the maintenance group and community perception about the maintenance industry. 2. The relative importance index was used to generate the model; X 1 is the number of respondents opting for “extremely important”; X2 is the number of respondents opting for “very important”; X3 is the number of respondents opting for “important”; X4 is the number of respondents opting for “somewhat important”; X 5 is the number of respondents opting for “not important. 3. Facilities managers of public hospitals rated the factors in this group to be of low level of importance while facilities managers of private hospitals rated the factors highly	- Uses log transform OLS model

Ferraris, V. A., Ferraris, S. P., & Singh, A. (1998). **Operative outcome and hospital cost.** The Journal of Thoracic and Cardiovascular Surgery, 115(3), 593–603. https://doi.org/10.1016/s0022-5223(98)70324-1	CONTEXT + METHODOLOGY	1. To facilitate analysis, the 32 different cost centers were grouped into seven different costs 2. Linear regression analysis was used to determine the correlation between hospital cost and LOS. 3. The Pearson correlation coefficient was calculated from the regression line, and an analysis of residuals was performed to determine the quality of the regression 4. A stepwise multivariate linear regression model was used to find independent predictors of increased hospital cost 5. Only preoperative variables were included in the multivariate models. Intraoperative or postoperative variables are tough to determine before the operation 6. Hospital cost was compared for each of three different patient outcomes, operative death, serious postoperative morbidity, and uncomplicated postoperative course.Hospital costs in uncomplicated cases averaged $31,579 (95% confidence interval [CI] $21,944 to $49,849). Costs in patients who had serious morbidity averaged $60,335 (95% CI $28,381 to $130,897), versus an average cost of $74,466 (95% CI $27,102 to $198,025) in patients who did not survive operation 7. analysis done for a) Hospital cost and outcome., b) LOS and outcome., c) Relationship between cost and hospital LOS, d) Multivariate analysis of risk factors for increased cost., e) Hospital cost and risk for operative death., f) Hospital cost and type of operation., d) Risks for death and increased LOS.	-

Links to Data Source & Dictionary

Dataset: https://data.cms.gov/provider-compliance/cost-report/hospital-provider-cost-report
Data Dictionary: https://data.cms.gov/resources/hospital-provider-cost-report-data-dictionary

R-Code

```
# Data: https://data.cms.gov/provider-compliance/cost-report/hospital-provider-cost-report
# Data Dictionary: https://data.cms.gov/resources/hospital-provider-cost-report-data-
dictionary
# Multi-level data: Upper level: State
#                   Lower level: Year
#
# Statistical Data Mining:  Factors Affecting Hospital Financial Stability
#   Durga Prasad Somarouthu
#   Md Akib Ali Sardar
#   Pruthvi Kokku
#   Sahil Shah
# Date : 2022-04-30

#Import the libraries
library(readr)
library(tidyverse)
library(dplyr)
library(ggplot2)
library(lme4)
library(car)

getwd()
# Read & preprocess data
df_14 <- read.csv("CSV_2014_Hospital_Cost_Report.csv")
df_14$year <- "2014"
df_15 <- read.csv("CSV_2015_Hospital_Cost_Report.csv")
df_15$year <- "2015"
df_16 <- read.csv("CSV_2016_Hospital_Cost_Report.csv")
df_16$year <- "2016"
df_17 <- read.csv("CSV_2017_Hospital_Cost_Report.csv")
df_17$year <- "2017"
df_18 <- read.csv("CSV_2018_Hospital_Cost_Report.csv")
df_18$year <- "2018"
df_combined <- bind_rows(df_14, df_15, df_16, df_17, df_18)
str(df_combined)
message("Total no of records = ", count(df_combined))


## rename the provider type and type of control to the meaningful names from data dictionary

df_combined$Provider.Type <-
  recode_factor(
    df_combined$Provider.Type,
    `1` = "General Short Term",
    `2` = "General Long Term",
    `3` = "Cancer",
    `4` = "Psychiatric",
    `5` = "Rehabilitation",
    `6` = "Religious Non-Medical Health Care Institution",
    `7` = "Children",
    `8` = "Alcohol and Drug",
    `9` = "Other"
  )
df_combined$Provider.Type <-
```

```
  relevel(df_combined$Provider.Type, "General Short Term")
df_combined$Type.of.Control <-
  recode_factor(
    df_combined$Type.of.Control,
    `1` = "Voluntary Nonprofit-Church",
    `2` = "Voluntary Nonprofit-Other",
    `3` = "Proprietary-Individual",
    `4` = "Proprietary-Corporation",
    `5` = "Proprietary-Partnership",
    `6` = "Proprietary-Other",
    `7` = "Governmental-Federal",
    `8` = "Governmental-City-County",
    `9` = "Governmental-County",
    `10` = "Governmental-State",
    `11` = "Governmental-Hospital District",
    `12` = "Governmental-City",
    `13` = "Governmental-Other"
  )

df_combined$Type.of.Control <-
  relevel(df_combined$Type.of.Control, "Proprietary-Individual")

## this is a cost report of all the hospitals with fiscal start and end date
## for our analysis we are considering the hospital that are reported with a
## range of 364 days->1 year i.e 90% of the dataset
df <- df_combined
df$noOfDays <-
  as.Date(df$Fiscal.Year.End.Date, format <-
            "%m/%d/%Y") - as.Date(df$Fiscal.Year.Begin.Date, format <-
                                    "%m/%d/%Y") == 364

df <- df[df$noOfDays == TRUE,]


message("Total no of records with fiscal range of 1 year = ", count(df))
#Handling Na values
#remove rows with na from y variable-Cost.To.Charge.Ratio
colSums(is.na(df))

df <- df[!is.na(df$Cost.To.Charge.Ratio),]

message("Total no of records removing na from Cost.To.Charge.Ratio = ",
        count(df))


## verify the dataset and remove invalid records. Here Cost.To.Charge.Ratio is calculated
## manually and compared with actual y variable. Mismatch records are removed
df$verifyC2C <-
  round((
    df$Total.Costs / df$Combined.Outpatient...Inpatient.Total.Charges
  ) - (df$Cost.To.Charge.Ratio),
  1
  )
df <- df[df$verifyC2C == 0,]
df <- df[df$Cost.To.Charge.Ratio < 100,]

message("Total no of records after verifying Cost.To.Charge.Ratio = ",
        count(df))


df <- df[!is.na(df$Total.Days..V...XVIII...XIX...Unknown.),]
```

```
df$Total.Days.Title.V <-
  ifelse(is.na(df$Total.Days.Title.V), 0, df$Total.Days.Title.V)

df <- df[!is.na(df$Total.Days.Title.XVIII),]

df <- df[!is.na(df$Total.Days.Title.XIX), ]

message("Total no of records removing na from both the titles = ", count(df))
## calculate unknown in total days
df$total_days_with_titles <-
  (df$Total.Days.Title.V) + (df$Total.Days.Title.XVIII) + (df$Total.Days.Title.XIX)
df$total_days_unknown <-
  (df$Total.Days..V...XVIII...XIX...Unknown.) - (df$total_days_with_titles)


df$total_days_unknown <-
  ifelse(df$total_days_unknown == 0, 1, df$total_days_unknown)



df$Total.Days.XVIII.medicare.ratio <-
  df$Total.Days.Title.XVIII / df$Total.Days..V...XVIII...XIX...Unknown.
df$Total.Days.XIX.medicaid.ratio <-
  df$Total.Days.Title.XIX / df$Total.Days..V...XVIII...XIX...Unknown.
df$Total.Days.V.ratio <-
  df$Total.Days.Title.V / df$Total.Days..V...XVIII...XIX...Unknown.
df$Total.Days.unknown.ratio <-
  df$total_days_unknown / df$Total.Days..V...XVIII...XIX...Unknown.

df <- df[!is.na(df$Total.Assets),]

message("Total no of records after removing na from totalassets = ",
        count(df))

df <- df[!is.na(df$Total.Current.Liabilities),]
df <- df[!is.na(df$Total.Long.Term.Liabilities),]

df <- df[!(df$Total.Current.Liabilities < 0),]
df <- df[!(df$Total.Long.Term.Liabilities < 0),]
message("Total no of records after removing na from liabiities = ", count(df))

df$debt.to.asset.ratio <-
  (df$Total.Current.Liabilities + df$Total.Long.Term.Liabilities) / df$Total.Assets

df <- df[df$debt.to.asset.ratio > 0, ]

message("Total no of records after cleaning debt.to.asset.ratio = ",
        count(df))

df <- df[df$Total.Income > 0, ]

df <- df[!is.na(df$Total.Income),]

message("Total no of records after cleaning Total.Income = ", count(df))


df <- df[df$Total.Unreimbursed.and.Uncompensated.Care > 0, ]

df <- df[!is.na(df$Total.Unreimbursed.and.Uncompensated.Care),]

message(
  "Total no of records after cleaning Total.Unreimbursed.and.Uncompensated.Care = ",
  count(df)
```

```
)


df$Type.of.Control <-
  relevel(df$Type.of.Control, "Proprietary-Individual")
## factor the char features
df$State.Code <- factor(df$State.Code)
df$State.Code <- relevel(df$State.Code, "FL")
df$Rural.Versus.Urban <-
  factor(df$Rural.Versus.Urban)
df$year <- factor(df$year)

df$Provider.Type <- as.factor(df$Provider.Type)
df$Type.of.Control <-
  as.factor(df$Type.of.Control)

## export the combined raw dataset
write.csv(df_combined, "hospitals_df_for_analysis.csv")

## export the dataset after cleaning
write.csv(df, "hospitals_df_for_analysis.csv")

## selective features from the predictor table
required_cols <- c(
  'rpt_rec_num',
  'Hospital.Name',
  'State.Code',
  'Rural.Versus.Urban',
  'Provider.Type',
  'Type.of.Control',
  'FTE...Employees.on.Payroll',
  'Total.Days.Title.V',
  'Total.Days.Title.XVIII',
  'Total.Days.Title.XIX',
  'Total.Days..V...XVIII...XIX...Unknown.',
  'Total.Days.XVIII.medicare.ratio',
  'Total.Days.XIX.medicaid.ratio',
  'Total.Days.V.ratio',
  'Total.Days.unknown.ratio',
  'Number.of.Beds',
  'Total.Bed.Days.Available',
  'Total.Discharges.Title.V',
  'Total.Discharges.Title.XVIII',
  'Total.Discharges.Title.XIX',
  'Total.Discharges..V...XVIII...XIX...Unknown.',
  'Cost.of.Charity.Care',
  'Total.Bad.Debt.Expense',
  'Cost.of.Uncompensated.Care',
  'Total.Unreimbursed.and.Uncompensated.Care',
  'Overhead.Non.Salary.Costs',
  'Depreciation.Cost',
  'Total.Costs',
  'Inpatient.Total.Charges',
  'Outpatient.Total.Charges',
  'Combined.Outpatient...Inpatient.Total.Charges',
  'Wage.Related.Costs..Core.',
  'Total.Salaries..adjusted.',
  'Cash.on.Hand.and.in.Banks',
  'Total.Current.Assets',
  'Total.fixed.Assets',
  'Total.Assets',
  'Total.Current.Liabilities',
```

```
  'Total.Long.Term.Liabilities',
  'Inpatient.Revenue',
  'Outpatient.Revenue',
  'Less.Total.Operating.Expense',
  'Total.Income',
  'Cost.To.Charge.Ratio',
  'debt.to.asset.ratio',
  'Net.Revenue.from.Medicaid',
  'year',
  'total_days_unknown',
  'Health.Information.Technology.Designated.Assets'
)

length(colnames(df))
# 138 features in total

length(required_cols)
# required features are 50

## creating a dataframe with the required columns for analysis - not all columns will be used
in modeling.
hospitals_df <- df[, c(required_cols)]
## prepare a temp dataset for correlation plot
hospitals_df_filtered_na <- na.omit(hospitals_df)
colSums(is.na(hospitals_df_filtered_na))
summary(hospitals_df_filtered_na)

correlation_cols <- c(
  'FTE...Employees.on.Payroll',
  'Number.of.Beds',
  'Total.Days.XVIII.medicare.ratio',
  'Total.Days.XIX.medicaid.ratio',
  'Total.Days.V.ratio',
  'Total.Days.unknown.ratio',
  'Total.Bed.Days.Available',
  'Total.Discharges.Title.V',
  'Total.Discharges.Title.XVIII',
  'Total.Discharges.Title.XIX',
  'Cost.of.Charity.Care',
  'Total.Bad.Debt.Expense',
  'Total.Unreimbursed.and.Uncompensated.Care',
  'Overhead.Non.Salary.Costs',
  'Depreciation.Cost',
  'Inpatient.Total.Charges',
  'Outpatient.Total.Charges',
  'Total.Salaries..adjusted.',
  'Cash.on.Hand.and.in.Banks',
  'Total.Current.Assets',
  'Total.fixed.Assets',
  'Total.Assets',
  'Inpatient.Revenue',
  'Outpatient.Revenue',
  'Less.Total.Operating.Expense',
  'Total.Income',
  'Cost.To.Charge.Ratio',
  'debt.to.asset.ratio',
  'Net.Revenue.from.Medicaid',
  'Health.Information.Technology.Designated.Assets'
)
corr_df <- hospitals_df_filtered_na[, correlation_cols]
correlation_output <- cor(corr_df)
write.csv(correlation_output, "hospital_report_correlation.csv")
```

```
revised_correlation_cols <- c(
  'Total.Days.XVIII.medicare.ratio',
  'Total.Days.XIX.medicaid.ratio',
  'Total.Days.unknown.ratio',
  'Total.Unreimbursed.and.Uncompensated.Care',
  'debt.to.asset.ratio',
  'Total.Income',
  'Health.Information.Technology.Designated.Assets'
)

corr_revised_df <-
  hospitals_df_filtered_na[, revised_correlation_cols]
correlation_revised_output <- cor(corr_revised_df)
write.csv(correlation_revised_output,
          "hospital_report_revised_correlation.csv")

##The base model was built to evaluate important factor variables,Äô
##effect on Cost to Charge Ratio. Log transform was applied to the
##target variable (Cost to Charge Ratio) to induce normality.
##All interpretations will reference percentage changes in Cost to Charge Ratio
##as a result of the predictors considered.

base_model <- lmer(
  log(Cost.To.Charge.Ratio) ~  Rural.Versus.Urban +
    Provider.Type +
    Type.of.Control +
    (1 | year) + (1 | State.Code),
  data = df,
  REML = FALSE
)
summary(base_model)
ranef(base_model)

##residual vs Fitted
plot(base_model)

##QQ plot - Normality test
qqnorm(resid(base_model)) # Q-Q plot
qqline(resid(base_model), col = "red")

## Homoscedasticity test
bartlett.test(list(resid(base_model), fitted(base_model)))

## multicollinearity test
vif(base_model)


##To evaluate the effect of Medicare and Medicaid on Cost to Charge
##Ratio we add the following derived variables from feaure engineering to the
##above base model: Total.Days.XIX.medicaid.ratio,
##Total.Days.XVIII.medicare.ratio, Total.Days.unknown.ratio,debt.to.asset.ratio
medicare_medicaid_model <- lmer(
  log(Cost.To.Charge.Ratio) ~  Rural.Versus.Urban +
    Provider.Type +
    Type.of.Control +
    log(Total.Days.XIX.medicaid.ratio) +
    log(Total.Days.XVIII.medicare.ratio) +
    log(Total.Days.unknown.ratio) +
    log(debt.to.asset.ratio) +
    (1 | year) + (1 | State.Code),
  data = df,
  REML = FALSE
)
```

```
summary(medicare_medicaid_model)
ranef(medicare_medicaid_model)


##residual vs Fitted
plot(medicare_medicaid_model)
##QQ plot - Normality test
qqnorm(resid(medicare_medicaid_model)) # Q-Q plot
qqline(resid(medicare_medicaid_model), col = "red")
## Homoscedasticity test
bartlett.test(list(
  resid(medicare_medicaid_model),
  fitted(medicare_medicaid_model)
))
## multicollinearity test
vif(medicare_medicaid_model)


##This model analyzes effects of predictors on Cost to Charge Ratio
#for the subset of hospitals which have invested in Health IT Assets.
#We needed to build a separate model


hit_df <-
  df[df$Health.Information.Technology.Designated.Assets > 0, ]
hit_df <-
  hit_df[!is.na(df$Health.Information.Technology.Designated.Assets),]
message(
  "Total no of records after cleaning Health.Information.Technology.Designated.Assets = ",
  count(hit_df)
)

hit_model <- lmer(
  log(Cost.To.Charge.Ratio) ~  Rural.Versus.Urban +
    Provider.Type +
    Type.of.Control +
    log(Total.Days.XIX.medicaid.ratio) +
    log(Total.Days.XVIII.medicare.ratio) +
    log(Total.Days.unknown.ratio) +
    log(Health.Information.Technology.Designated.Assets) +
    log(debt.to.asset.ratio) +
    (1 | year) + (1 | State.Code),
  data = hit_df,
  REML = FALSE
)
summary(hit_model)
vif(hit_model)
ranef(hit_model)

##residual vs Fitted
plot(hit_model)
##QQ plot - Normality test
qqnorm(resid(hit_model)) # Q-Q plot
qqline(resid(hit_model), col = "red")
## Homoscedasticity test
bartlett.test(list(resid(hit_model), fitted(hit_model)))
## multicollinearity test
vif(hit_model)

library("stargazer")
stargazer(base_model, medicare_medicaid_model, model_hit, type = "text",
  single.row = TRUE
```