June 27, 2024

Robert Patton

Western GOVERNORS’ University

Rpatt33@wgu.edu

Executive Summary

D214

Table of Contents

1. **Problem and Hypothesis**2
2. **Summary of Data Analysis Process**2
3. **Outline of Findings**3
4. **Study Limitations**3
5. **Proposed Actions**3
6. **Study Benefits** …………………………………………………………………………3-4
7. **Sources**4

**Executive Summary and Implications**

1. **Problem and Hypothesis**

**Research Question**: To what extent do the county, OSHPDID, and hospital system variables affect the number of adverse events?

**Hypothesis:** **Null hypothesis-** The county, OSHPDID, and hospital system variables do not have statistical significance in the number of adverse events occurring.

**Alternate Hypothesis-** The county, OSHPDID, and hospital system variables have statistical significance affecting the number of adverse events.

Hospital ratings are based on risk-adjusted rates for California hospitals between the years of 2016 and 2021. Adverse events are recorded, based on specific performance measures for medical conditions, and are compared to total cases. To evaluate the risk of an adverse event occurring, a regression model can be used to look at a correlation between hospital location and its impact on adverse event occurrence.

1. **Summary of Data Analysis Process**

The data information was examined and reviewed, using Python, for variable selection relative to the research question. County, OSHPDID, # of adverse events, and hospital system were selected for data analysis. These variables became the sole focus of study in a multiple linear regression model, where the correlation between independent and dependent variables could be evaluated. Exploratory data analysis helped reveal the most appropriate variables to be used in the regression model and allowed me to transform these variables into values necessary for data analysis. Independent variables included county, OSHPDID, and hospital system and the dependent variable included # of adverse events.

One-hot encoding, by creating dummy variables, was used to appropriately set up the data for use in a multiple linear regression model. Missing variables were removed from the data set because the hospital either did not report any information on cases or cases equated to less than three. An initial regression model was created and evaluated for performance using the R-squared metric. Insignificant variables were removed based on a p-value greater than 0.05, using a backward stepwise elimination method, where a new regression model was created after each insignificant variable was removed. Once all variables had a p-value less than or equal to 0.05, a final regression model was created and evaluated again on the R-squared metric. A final regression model produced an R-squared value of 0.90 and the residual standard of error was also evaluated for model validation.

1. **Outline of Findings**

As mentioned, the final regression model produced an R-squared value of 90%. This establishes that the model is highly accurate, and the alternate hypothesis can be accepted for statistical significance between the independent and dependent variables of the study. Furthermore, to validate the model’s accuracy even further predicted values and residuals were created, plotted in a scatterplot, and the residual standard of error was calculated. A residual standard of error closer to zero indicates a better model, where a result of 11.38 for this model validates its accuracy.

1. **Study Limitations**

The biggest limitation of this analysis is that it only includes four years’ worth of data, and some of the data had to be removed because either the hospitals did not report any cases, or the values were less than 3. However, even if cases were less than 3, it could still be of value to the study and produce very different results had those been included. If there were ten years’ worth of data for this project, the results would probably be a lot more accurate and there would probably be data available for each of those variables where values were dropped.

1. **Proposed Actions**

Future studies on this matter can be improved in many ways, however. I propose that one method for improving this study would be to look at ambulance transport times for those patients experiencing the performance measures/medical conditions listed in this data set. Adding that metric to this data set would most certainly help explain a lot of the adverse event outcomes. Communities, where hospitals have negative ratings and higher adverse event rates, may see some correlation between what happens before patients arrive at the hospitals, via emergency vehicles, and the measures taken in the hospital. Another measure that could improve this study would be to combine some of the performance measures. For example, there are four-stroke categories in this study: Acute Stroke Subarachnoid, Acute Stroke Hemorrhagic, Acute Stroke, and Acute Stroke Ischemic. Subarachnoid and hemorrhagic strokes could be combined into one category as both infer that a patient had a “bleeding brain stroke”. Whereas an acute stroke and acute ischemic stroke infer a “blood clot brain stroke”. By combining the categories, the study can be simplified into the two main types of strokes and research could go into identifying what is the root cause of adverse events related to hemorrhagic and ischemic strokes, one of the most debilitating and life-threatening medical conditions.

1. **Study Benefits**

Hospital rankings are generally produced by entities such as the Centers for Medicare and Medicaid Services (CMS), evaluating quality measures for certain metrics such as emergency room wait times or patient care based on certain medical conditions (DeAngelis 2016). This analysis project could benefit California hospitals by improving their hospital ratings by improving metrics on adverse events. For example, a patient having a stroke medical emergency may not live close to a hospital with neurological services, thus increasing the likelihood of an adverse event at that hospital. Using regression analysis to determine if location influences adverse events can help improve hospital ratings by showing executives which medical specialties may be beneficial to add, or it can lead hospital executives to explain specific recurrent adverse events based on a specific patient population type in that location, such as elderly patients who are more prone to having hip fractures. The main goal of this data analysis project was to determine if there is a correlation between hospital demographic variables and adverse events. Furthermore, if such is true, hospital ratings may be impacted by this finding, and understanding this correlation may help to improve a hospital rating, which improves business operations. For example, in 2017 imposed fines by the CMS for hospital readmissions equated to over half a billion dollars (Upadhyay et al. 2019). An increase in hospital admissions, due to adverse events, will impact revenue and produce unexpected expenses in the form of fines. Overall, this study is not only beneficial for business operations but also for the local community members in which each hospital serves. This is a business of improving and saving human life, and to understand and improve the risk of adverse events for our California hospitals is to improve the quality of human life and confidence in our state’s healthcare system.

1. **Sources**

DeAngelis, C. D. (2016, December). How helpful are hospital rankings and ratings for the public’s health? The Milbank quarterly. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5192961/>

Upadhyay, S., Stephenson, A. L., & Smith, D. G. (2019). Readmission Rates and Their Impact on Hospital Financial Performance: A Study of Washington Hospitals. Inquiry: a journal of medical care organization, provision, and financing, 56, 46958019860386. <https://doi.org/10.1177/0046958019860386>