A Regression Analysis of Patient and Hospital Characteristics Influencing Length of Stay in South Carolina Hospitals

Michael A. Schlenk, Kennesaw State University

Faculty Advisor: MinJae Woo, PhD

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

ABSTRACT

The goal of this project was to examine various factors affecting a patient's length of stay (LOS) in the hospital. A total of 397,846 patients from Prisma Health hospitals in South Carolina were retrieved. Three different regression models were built to determine the best predictors of LOS. The number of diagnoses, age at visit, admission to the emergency department, hospital, number of procedures, occupancy rate at the start of the visit, and paying out of pocket were significant predictors for patients' LOS while race, sex, and marital status were insignificant predictors. Future research should further examine ways for hospitals to promote efficiency and patient management in order to decrease length of stay.

INTRODUCTION

Length of stay is an important indicator of a hospital's management and efficiency. The average LOS can provide insight to the quality of treatment, availability of timely care, patient experience, and hospital profit. With increased LOS, more time is spent on bed management with decreased time to treat new patients. Shorter LOS may be associated with decreased likelihood of infections, medication side effects, and improved treatment. As such, the purpose of this study is to determine which factors are associated with the length of hospital stay at 9 major hospitals in South Carolina.

METHODS

Patient records were retrieved from a centralized database of patients admitted to one of 9 hospitals affiliated with Prisma Health in upstate South Carolina between January 2014 and December 2016. A total of 397,846 individual patients with at least one day of hospital stay were identified. The following 14 variables have been selected for the study: (1) Age at the time of Visit , (2) Admission Source, (3) Financial Class Group (Insurance Type), (4) Hospital, (5) Length of Hospital Stay, (6) Sex, (7) Marital Status, (8) Race, (9) Number of Diagnoses, (10) Number of Procedures, (11) Number of Hospital Patients at the Start of the Visit, (12) Number of Hospital Patients at the and End of the Visit, (13) Max Occupancy of Hospital at the Start of the Visit, (14) Max Occupancy of Hospital at the End of the Visit

Building the Best Model
1.) Creating New Variables:
- "NUM_AGE_AT_VISIT": discretized the "AGE_AT_VISIT" variable
- "OCCUPANCY_RATE": divided the max occupancy at the start of the visit by the number of patients at the start of the visit.

2.) Lasso's Regression was run to find the variables with the most influence
- Female Sex, Number of Procedures, Number of Diagnoses, Number of Hospital Patients at the End of the Visit

3.) Simple linear regressions of all variables were run find the lowest MSE

Model 1: built by adding 1 variable at a time until the lowest MSE was obtained
Model 2: built by adding 1 variable at a time to the 4 most influential variables until the lowest MSE was obtained.
Model 3: edited Model 2 by adding "OCCUPANCY_RATE", ADMISSION_SOURCE_ED, and FINANCIAL_CLASS_GROUP_SELF as well as removing Number of Hospital Patients at the End of the Visit.

Selection Criteria: lowest AIC, BIC, and p-value as well as highest Adjusted R-Squared

RESULTS

Number of Diagnoses: For every 1-increase in the number of diagnoses, length of stay (LOS) increased by 0.3675375 days on average given that all other variables are held constant. This association is significant.

Age at Visit: For every 1-year increase in a patient's age at visit, LOS decreased by -0.0124214 days given that all other variables are held constant. This association is statistically significant.

Admission to the Emergency Department (ED): As compared to the ED, all other admission sources were associated with a decreased LOS by -0.5316635 days given that all other variables are held constant. This association is statistically significant.

Hospital: The LOS compared by each hospital is statistically significant given all other variables are held constant. For example, Hospital 7 associated with an increase in LOS by 10.3891344 days on average.

Number of Procedures: For every 1 increase in the number of procedures, LOS increased by 2.8689386 days on average. This association is statistically significant.

Patient Occupancy Rate at Start of Visit: For every 1 increase in the occupancy rate at the start of visit, LOS decreased by -0.3197595 days on average. This association is statistically significant.

Self Payment: As compared to patients who paid out of pocket, having insurance is associated a decrease in LOS by -0.1826473 days on average. This association is statistically significant.

Race: A patient's race was not a significant predictor for their LOS.

Sex: A patient's sex was not a significant predictor for their LOS.

Marital Status: A patient's martial status was not a significant predictor for the LOS.

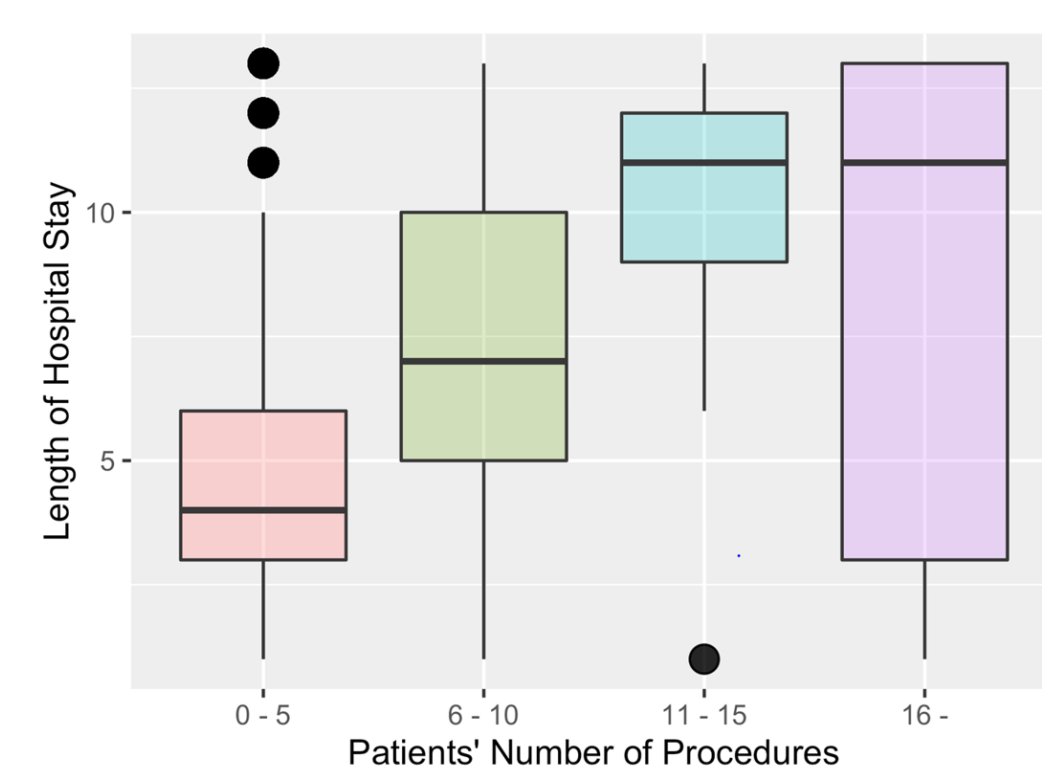Figure 1: LOS by Number of Procedures
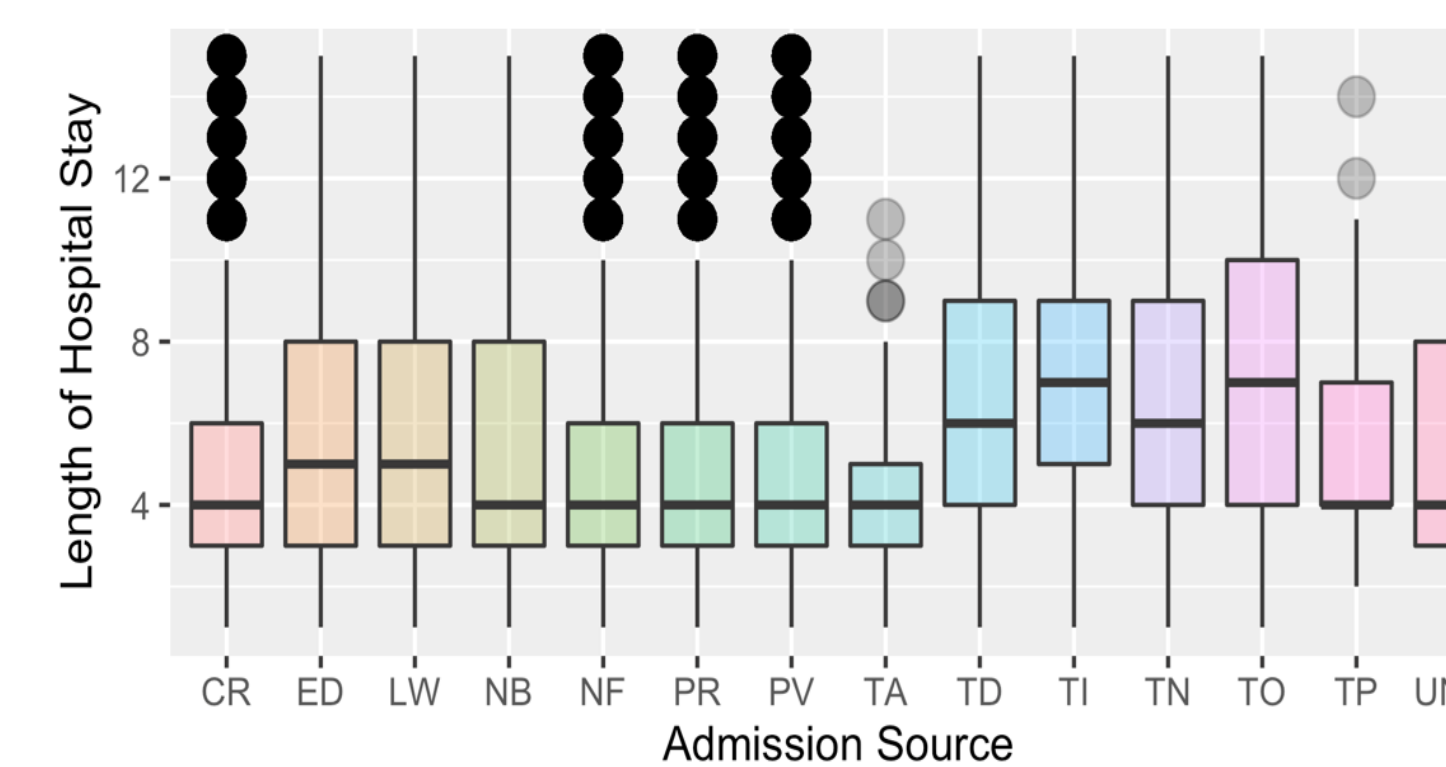Figure 2: LOS by Admission Source
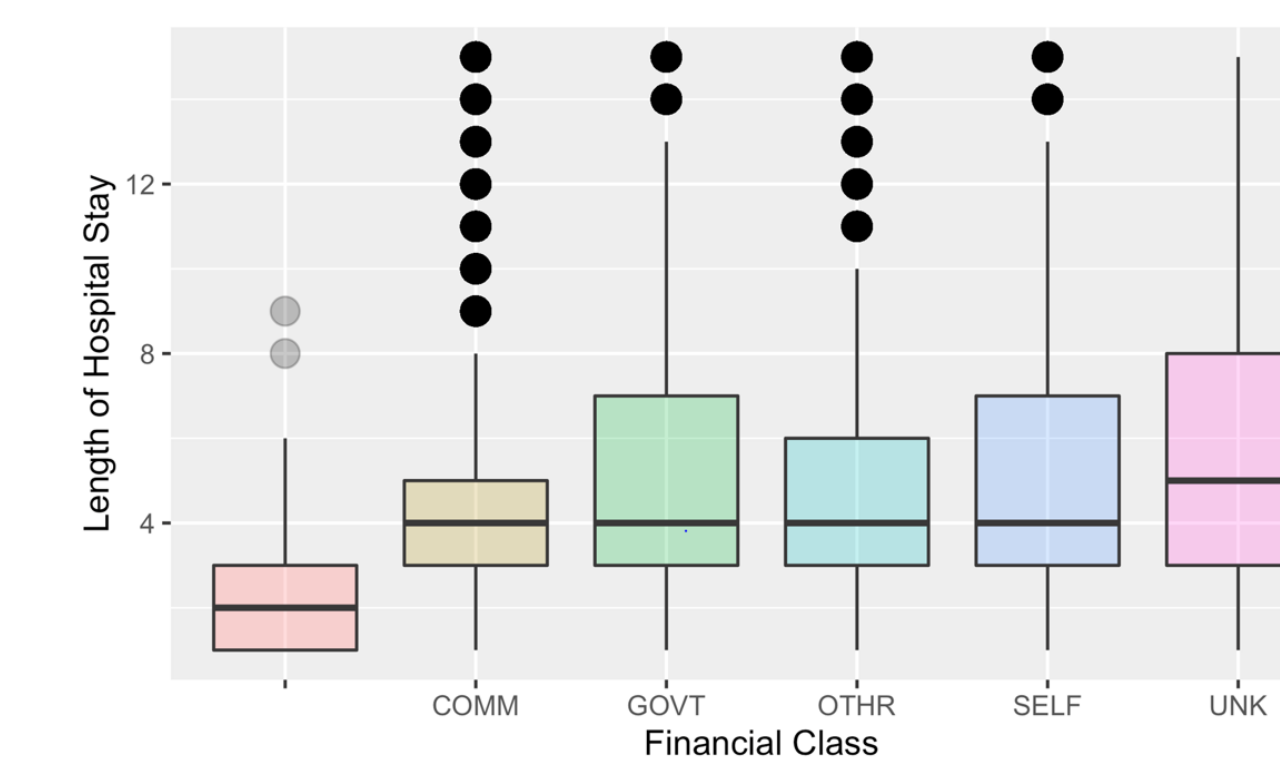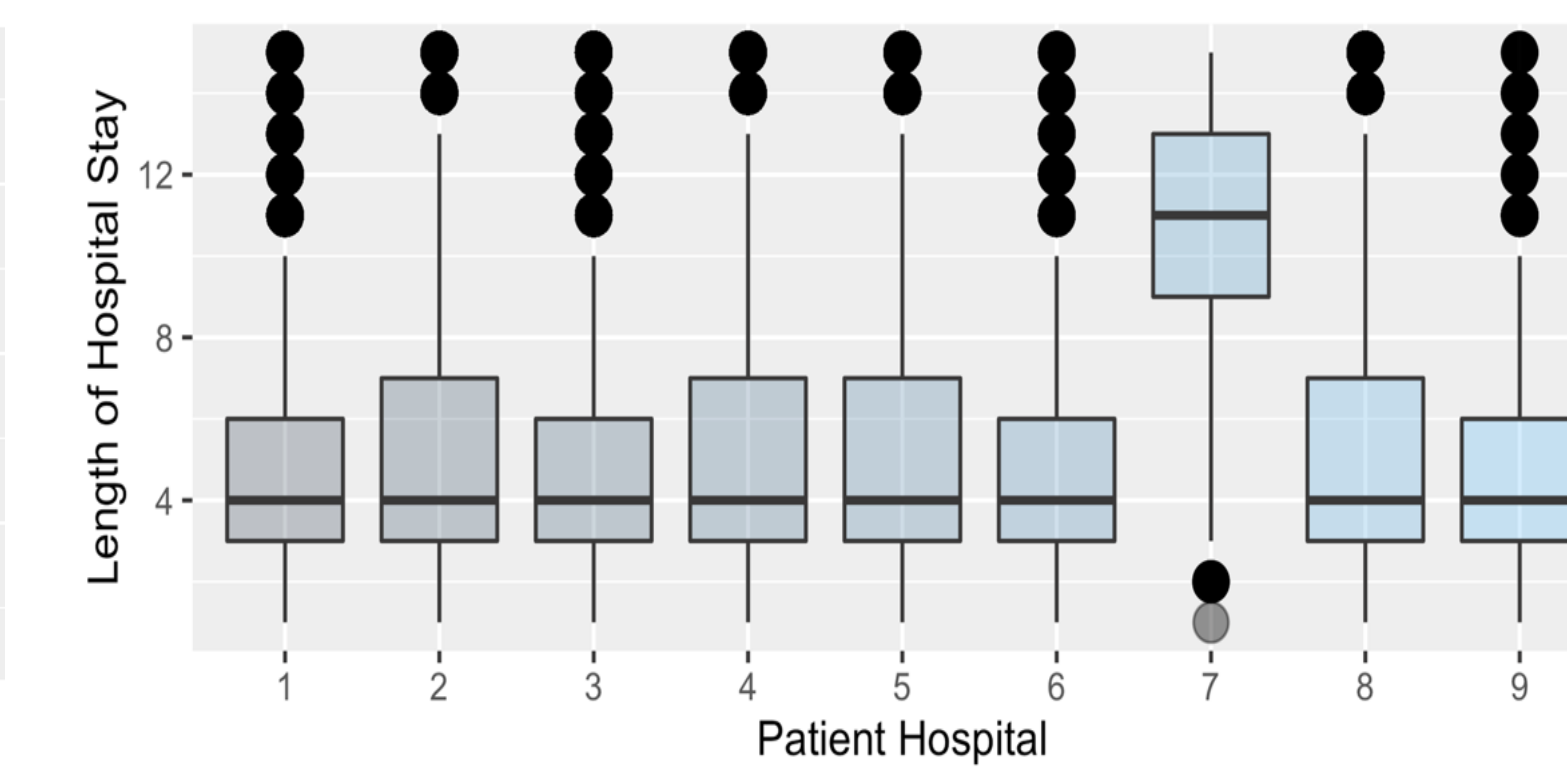Figure 3: LOS by Financial Class
Figure 4: LOS by Hospital

CONCLUSIONS

- The insignificant prediction of LOS by patient race, sex, and marital status are beneficial as it indicates a lack of discrimination in the hospitals examined.
- The differences in LOS for the 9 hospitals are likely attributed to other variables not examined in this study such as hospital funding, physician specialties, patient populations, type of insurance accepted, number of beds available, and more.
- The association between the number of diagnoses and procedures with an increased LOS does not account for the urgency. More urgent diagnoses and procedures likely increase LOS for those receiving them and thereby increase the LOS for patients with less urgent needs.
- Patients paying out of pocket may have shorter LOS due to the potential for hospitals to not receive payments on the treatments provided.
- The association between increased LOS and admission to the ED may be due to more severe injuries that require more diagnoses and/or procedures as well as receiving non-emergency patients.
- The age at visit may be associated with decreased LOS due agreeableness with physicians as well as the increased risk for infections and impairments on daily life.
- The occupancy rate's association with a decreased stay may indicate that hospitals are trying to keep a steady flow of patient overturn to maintain treatment quality and timely care. This variable is not common in the literature and should be further examined.
- Future research should continue to examine methods to decrease LOS utilizing patient information and occupancy in conjunction with hospital policies and management.

R CODE

R Packages
- mass: run the Akaike Information Criterion (AIC) stepwise selection to compare each of the models built
- glmnet: run Lasso's Regression to find the variables with the most influence on the model
- statar: explore the data frame with summarize, tabulate and percentile functions.
- ggplot2, tidyverse, hrbrthemes: create graphs with details of aesthetics, variable mapping, scales, and themes
- fastdummies: create dummy categorical variables to be added in the model