

Project Proposal

Nihar Sidhu (ns65) and Nehal Rawat (nr338)

September 21, 2017

1 Predicting Hospital Charges in New York State

Dataset Information:

We are interested in looking at the The Statewide Planning and Research Cooperative System (SPARCS) Hospital Inpatient Discharges dataset. This dataset has information about patients that visit hospitals across New York state. Features of the dataset include counties in New York State, demographics such as age and race of patients, length of stay of a patient at the hospital, patient's illness severity, patient's risk of mortality, payment types, total costs, and total charges.

Proposed Question:

We are interested in answering some of these questions through our data analysis: Can we predict which factors have the largest influence on a hospital charges? Given a patient's illness severity, length of stay at a hospital, and the hospital he/she is going to, can we predict what the total cost would be?

We can also use the algorithms we have learned in class to classify which counties and hospitals in New York State seem to be more expensive on average, and thus create a visualization of some sort. One concern we have about our data is that it is from 2012, and thus might not be as applicable today.

This data analysis is important because it will help patients understand how their medical costs came to be. The analysis can also help patients to choose hospitals that may have lower costs.