

Project Report

By: Priyanka Joshi

Analyze the Healthcare cost and Utilization in Wisconsin hospitals

Business Scenario

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on the healthcare costs and their utilization.

Expectations or goals:

The goals of this project are:

1. To record the patient statistics, the agency wants to find the age category of people who frequent the hospital and has the maximum expenditure.
2. In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis related group that has maximum hospitalization and expenditure.
3. To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.
4. To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for proper allocation of resources.
5. Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.
6. To perform a complete analysis, the agency wants to find the variable that mainly affects the hospital costs.

Analysis:

1. From all the age groups infants have the maximum visits to the hospital and has the maximum expenditure.
2. From the results we can see that the category 640 has the maximum entries of hospitalization and also has the highest total hospitalization cost (437978).
3. We can infer that p value is very high around 94% this means that we can reject the null hypothesis also race of the patients is not related to the hospitalization costs we can observe that the race 1 have 484 patients which is higher than other groups so data is skewed and it will affect the results by ANOVA
4. As we can see from the results that p value for age is very less this means it is an important factor in the hospital costs as seen by the significance levels and p-values. gender has also less p value means it is also having the impact on cost and same with intercept
5. By observing the results of model, we can infer that except for the intercept p values are very high.

The very high p-value signifies that there is no linear relationship between the given variables.

That is, with just the age, gender, and race, it is not possible to predict the LOS (length of stay) of a patient

6. From the output we can say that Age and LOS (Length of stay) affects the hospital costs as higher length of stay of the patients will result in higher hospital costs