Name : Swapnil Narwade

swapnil.narwade3@gmail.com

Healthcare Cost Analysis Report

Background and Objective :

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 healthcare costs and their utilization.

Domain : Healthcare

Dataset Description :

Attribute	Description
Age	Age of the patient discharged
Female	A binary variable that indicates if the patient is female
Los	Length of stay in days
Race	Race of the patient (specified numerically)
Totch	Hospital discharge costs
Aprdrg	All Patient Refined Diagnosis Related Groups

Name : Swapnil Narwade

swapnil.narwade3@gmail.com

Healthcare Cost Analysis Report

Analysis to be done :

1. To record the patient statistics, the agency wants to find the age category of people who frequently visit 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 the 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 hospital costs.

Name : Swapnil Narwade

swapnil.narwade3@gmail.com

Healthcare Cost Analysis Report

Conclusions :

1. To record the patient statistics, the agency wants to find the age category of people who frequently visit the hospital and has the maximum expenditure.

- **As per the analysis, age group 0 to 1 visits the hospital most frequently.**
- **Also their maximum expenditure is 678,118$ /-**

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.

- **640 diagnosis related group had a max cost of 437,978$ /-**

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.

- **We get correlation p-value 0.69 in our statistical performing in R; which is much higher than 0.05**
- **If p-value is greater than 0.05 mean we reject the null hypothesis, hence there is no significant relationship.**
- **Meaning, we can say that the race of the patient does not have effect on hospitalization cost, hence there is no malpractice as per the statistical analysis.**

4. To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for the proper allocation of resources.

- **Since the p-values of Age is much lesser than 0.05, it means Age has the most statistical significance.**
- **Similarly, gender is also less than 0.05.**
- **Hence, we can conclude that the model is statistically significant.**

Name : Swapnil Narwade

swapnil.narwade3@gmail.com

Healthcare Cost Analysis Report

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.

- **The p-value is higher than 0.05 for age, gender and race, indicating there is no linear relationship between these variables and length of stay.**
- **Hence, age, gender and race cannot be used to predict the length of stay of inpatients.**

6. To perform a complete analysis, the agency wants to find the variable that mainly affects hospital costs.

- **We can say that Age, Length of stay (LOS) and patient refined diagnosis related groups (APRDRG) have three stars (***) next to it. So they are the ones with statistical significance.**
- **These three variables mainly effects the hospital cost.**

Thank you !