HEALTHCARE COST ANALYSYS

Domain: Healthcare

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.

Analysis to be done:

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 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.

Approach:

1. This can be found out by applying summary() function on AGE and aggregate() function on TOTCHG summing expenditure of all ages, and extracting the max from both and comparing them.

2. This can be found out by applying summary() function on APRDRG and aggregate() function on TOTCHG summing expenditure of all diagnostic groups, and extracting the max from both and comparing them.

3. ANOVA function can be applied over the variables TOTCHG (independent) & RACE (dependent), using this we can determine the relationship between the treatment of a patient and their race.

4. Linear Regression Model can be applied over the variables TOTCHG (independent) & AGE, FEMALE(dependent), using this we can determine the severity of expenditure to allocate resources efficiently.

5. Linear Regression Model can be applied over the variables LOS (independent) & AGE, FEMALE, RACE (dependent), using this we can predict the length of stay using the 3 dependent variables.

6. Linear Regression Model can be applied over the variables TOTCHG (independent) & AGE, FEMALE, RACE, LOS, APRDRG (dependent), using this we can analyze all the variables which affect the total expenditure.