Healthcare_project

By:Krishna Aryal
Trainer:Deepti

Data Understanding

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.

Data understanding

Variables:

1.Age Age of the patient discharged,

2. Female A binary variable that indicates if the patient is female

3. Los Length of stay in days

4. Race

5. Totchg Hospital discharge costs

6. Aprdrg All Patient Refined Diagnosis Related Groups

Model used

Anova test: to figure out malpractices so that hospital costs affected by race group

Multiple linear regression: to figure out health cost with other variables.

Codes: for anova

ano_model <- aov(TOTCHG~factor_race,data=hos_cost2)

summary(ano_model)

For linear regression: lm_model <- lm(hos_cost$TOTCHG ~ hos_cost$AGE+hos_cost$FEMALE)

summary(lm_model)

Model Interpretation and findings

- To record patient statistics and finding age category who visited mostly histogram is plotted. Expenditure is calculated using aggregate function.
- Severity of diagnosis related group 640 diagnosis group has most hospitalization and expenditure.
- It is also found that there is no malpractice for race related group.
- Hospital cost is related with age and gender wise.
- To determine length of stay related with age,gender and race linear regression model is used.
- Lastly using multiple regression model hospital cost affecting factors found. Length of stay, diagnosis group and age.