**Healthcare cost analysis**

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 analyse 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 patients is female

Los Length of stays in days

Race Race of patient (specified numerically)

Totchg Hospital discharge costs

Aprdrg All Patient Refined Diagnosis Related Groups

**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.

**Solution:** From figure1 and Figure2 below, Patients in the age category of 0-5 visited the hospital more and have the highest expenditure than every other category.

Chart, bar chart

Description automatically generated

Figure

Chart, bar chart

Description automatically generated

Figure

**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.

**Solution:** From figure3 and figure4 below, Patients in the diagnostic group 640, have the maximum hospitalization and expenditure than every other diagnostic group.

Chart, bar chart, histogram

Description automatically generated

Figure

Chart, bar chart, histogram

Description automatically generated

Figure

**3.** To make sure that there is no malpractice, the agency needs to analyse if the race of the patient is related to the hospitalization costs.

**Solution:** After calculating the Analysis of variance, it is observed that there is no relationship between the race and hospitalization cost as the p-values was greater than 0.05. We can also observe from figure5 that Race 1 has the highest expenditure and this is because they represent a larger percent of the patient in the hospital.

Graphical user interface

Description automatically generated with low confidence

Figure

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

**Solution:** Patients of Age 0 has the highest expenditure when grouping by age while the male patients have the highest expenditure when grouping by gender.

Chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated

**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.

**Solution:** From the result of the regression model, the Length of stay cannot be predicted using the age, gender, and race. The is no linear relationship between the variables and the result of p-value which is greater than 0.05 shows they are not statistically significant in the model.

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

**Solution:** Age, Length of stay, and Patients diagnostic related group are strong predictors of the hospital cost, with very low P-value. The linear model explains 53% of the variance of the dependent variable from the independent variables. We can generally conclude that there is a linear relationship between the dependent variable and the predictors.