

Project Proposal – Group 7

This project is based on observing the trends of the Annual Medical Cost associated with the Primary Insurance Holder, based on multiple factors such as their age, sex, BMI, No. of Children, Smoking tendency and their residential region.

Type of Study:

Observational Study.

Source of Data:

The data is taken from Kaggle Datasets.

<https://www.kaggle.com/mirichoi0218/insurance>

Variables:

Continuous Predictors:

- Age: Age of Primary Insurance Holder.
- BMI: Body Mass Index, henceforth referred to as BMI, is the ratio of the weight of an individual to the square of their height. The units are in Metric system – (kg / m^2).
- Children: Number of children covered by health insurance.
- Charges: Primary Insurance Holder medical costs charged by health insurance.

Categorical Predictors:

- Sex: Primary Insurance Holder Gender.
- Smoker: If the Primary Insurance Holder is a smoker or Non-Smoker.
- Region: The residential area of the Primary Insurance Holder in the US with 4 levels – northeast, southeast, southwest, northwest.

Response Variable:

Charges is the response variable in our project. We would like to determine if there is a linear relationship between the medical charges with age, BMI and No. of Children.

Hypothesis Test 1:

As the age increases, the charges of the primary insurance holder increase. The medical charges of the higher aged Primary Insurance Holders will be higher than Lower aged Primary Insurance Holders.

Hypothesis Test 2:

The medical charges increase with increase in No. of Children. The Primary Insurance Holders with a greater number of children have higher charges than the Insurance Holders with lesser number of children.

Hypothesis Test 3:

The Primary Insurance Holders with higher BMI have higher medical costs than the Primary Insurance Holder with lower BMI.