Regression Analysis of Medical Insurance Costs
Project Proposal

Project Statement

Medical insurance costs vary from person to person due to many differences in individuals. These differences include (and are not limited to) an individual's age, sex, smoking status, or BMI. These variables each effect the insurance cost to different extents, and it is hypothesised that the smoking status of an individual is a powerful predictor of medical insurance costs.

Deliverables

A dataset found through the online site Kaggle[1] lists 1338 instances of 7 variables (age, sex, smoking status, BMI, children, region, charges) with the target variable being `charges`, with aims to be utilised in a regression analysis to model and predict an individual's medical insurance costs.

Results of the analysis are to be presented in a report which will cover the following:

- Data handling and cleaning to ensure the information is consistent
- Implementing a regression model
- Analysis of the regression model including model adequacy checks
- Predictions of medical insurance costs
- Conclusions and summaries of the project.

Group Information

The project is to be completed individually by:

Name: Abhishekh Shankar
Student Num: s3652116

[1] https://www.kaggle.com/mirichoi0218/insurance