Regression Answers

1. The dataset was on a state, medical condition, year level. The training set had 1644 observations. These observations represented aggregated Medicare data.
2. The years included were 2014-2017.
3. The conditions included were arthritis, asthma, atrial fibrillation, chronic kidney disease, depression, diabetes, hyperlipidemia, hypertension, ischemia, psychotic disorder, pulmonary disease, and stroke.
4. For the regression model we used demographic information from the prior year, factor variables for the type of condition, and the prior year’s previous median price, to predict the next year’s median price for each condition.
5. The R-square of the model was .941.
6. The most important variable was the prior year’s cost, followed by the type of medical condition.
7. We trained the model on 70% of the data. The RMSE was $6,222. The MAE was $3,554.
8. We predicted 2018 data using observed data from 2017. A prediction interval was calculated using the standard error of the prediction. <https://newonlinecourses.science.psu.edu/stat462/node/151/>