

Assignment 4

Your Name Here

7/6/2021

These exercises will require you to load the `area_data.Rds` file.

1. Run a model (on a *training* subset of the data) that predicts the percent of the population in the labor force (dependent variable) as a function of the percent of the population that is insured (independent variable). Make sure you use the workflow process and present the regression table of results.
2. Summarize the coefficient/estimate for percent of the population that is insured in a sentence or two.
3. Calculate the model fit by calculating the rmse in the *testing* data.
4. Add variables for the census division and the percent of the population with commutes above 30 minutes. Comment on the estimates for both of these variables. (*N.B: the census division is a categorical variable*).
5. Calculate the model fit by examining the rmse in the testing data. Comment in a sentence on what the rmse means and how it compares to the rmse in step 3.
6. Create another model by adding at least two other variables (KEEP the same dependent variable).

Answer the following questions in a sentence for each: 6a. What is the rmse from your new model? 6b. How does it compare to the rmse from the previous model? 6c. Which predictors appear to be related to the outcome? How do you know?