**Data Exploration of Lung Disease Data**

Follow steps below to model the dataset for lung disease. The dataset is contained in the Canvas Folder for this assignment. Your answers to questions may be in the form of comments in your code.

1. Read into a data frame the file lungDisease.csv and store in variable lung\_data.

Note that this file contains a header. done

1. Fit a linear main effects model named fit\_main: variable rvtlc as a function of the main effects of the rest of the variables (age, yrs\_smoke, and yrs\_tot\_smoke). Comment on your model, including fitted coefficients, goodness of fit, and significance of the model as a whole. Explain the effect of each predictor variable on the response variable. (Note that RV/TLC is a measure of chronic obstructive pulmonary disease (COPD), where a larger ratio indicates obstruction, COPD.) done
2. Fit a linear model of named fit\_smoke of variable rvtlc as a function of predictor variable yrs\_smoke. Account for the difference in how fit\_smoke and fit\_main represent the effect of yrs\_smoke on rvtlc. done
3. Fit a full linear model named fit\_full (include main effects and 2nd order interactions) with variable rvtlc as the response variable the rest of the variables as the predictor variables. Compare this model to fit\_main. done
4. Beginning with fit-full, remove insignificant variables one-by-one to improve your model. Provide the equation of the fitted model. Explain how a 1-unit increase in each predictor variable affects the response variable. done
5. Add a column to your data frame to hold a boolean variable named has\_disease, which will be determined by variable rvtlc. If rvtlc >= 35 the subject is assigned has\_disease = 1, otherwise has\_disease = 0. done
6. Create a histogram for each variable, colored by has\_disease. Briefly comment on each histogram. done
7. Fit a logistic regression model with the variable has\_disease as a function of predictor variables age, yrs\_smoke, and yrs\_tot\_smoke. Contrast the results of this model to your linear model in #2.
8. Using the model of #4 as a starting model, use the step function to improve the model.
9. Fit a logistic regression model with has\_disease as a function of rvtlc. Briefly explain your results.