**R Module 9 Assignment – Multiple Linear Regression**

Consider the FEV dataset we looked at last module. We are interested in answering the question, “How does smoking status affect FEV?”

1. We will first consider a simple linear model with FEV as the outcome, and smoking status as the predictor. Fit the simple linear model, and summarize the results. (you can assume the model assumptions are met for now). Does your result make sense?
2. It is possible the relationship we observed in (1) is confounded by some other variables. Now, we will fit a multiple linear model to see if this is the case. We are trying to answer the question, “What is the relationship between smoking status and FEV?”
   1. We will consider the outcome to be FEV, and the potential predictors/confounders to be age, height, sex, and smoking status. Plot the potential predictors/confounders versus the outcome, FEV. Considering each graph, does there appear to be a linear relationship with FEV? Paste your plots and comments for the answer to (a).
   2. Fit a model with FEV as the outcome, and age, height, sex, and smoking status as the predictors. Do not use any interaction terms. Plot a histogram of the residuals and the normal probability plot of the residuals. Are the residuals approximately normally distributed? (Include the plots and your comments for your answer)
   3. Interpret the parameter estimates in the model (write out the equation with the parameters). How has the parameter estimate for smoking status changed from (a)? Why might it have changed?
   4. Use the model in (b) to answer the original question posed in the assignment. “What is the relationship between smoking status and FEV?”