STAT 308 – In Class Exercise 2

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you would not be given full credit on the exam.

The `state.x77` data matrix is automatically available in base `R`. The dataset provides information on information from all 50 US states from the 1970s about the following categories:

- `Population`: population estimate as of July 1, 1975
- `Income`: per capita income (1974)
- `Illiteracy`: illiteracy (1970, percent of population)
- `Life Exp`: life expectancy in years (1969-71)
- `Murder`: murder and non-negligent manslaughter rate per 100,000 population (1976)
- `HS Grad`: percent high-school graduates (1970)
- `Frost`: mean number of days with minimum temperature below freezing (1931–1960) in capital or large city
- `Area`: land area in square miles

1. Suppose we wish to know if state's illiteracy rate is related linearly to at least one of the other variables in the dataset.

a. State the least squares regression line with illiteracy as the response variable with all the other variables as explanatory variables.

```
state.x77 <- as.data.frame(state.x77)
```

b. Report and interpret the r^2 value for the full linear model in context of the problem.

c. Perform a formal hypothesis test to answer if illiteracy rate is related linearly to at least one other variable in the dataset. Please report all the information needed to perform a hypothesis test. Assume $\alpha = 0.05$.

2. Suppose we know that murder rate is related linearly to life expectancy. Perform a hypothesis test to determine if adding illiteracy rate to our linear model for life expectancy with murder rate included significantly improves the predictive ability of our model. Assume $\alpha = 0.05$.

3. Suppose we know that murder rate is related linearly to life expectancy. Perform a hypothesis test to determine if adding high school graduation rate to our linear model for life expectancy with murder rate included significantly improves the predictive ability of our model. Assume $\alpha = 0.05$.

4. Perform a hypothesis test to determine if adding all the addtional variables to our linear model for life expectancy with both murder rate and high school graduation rate already included significantly improves the predictive ability of our model. Assume $\alpha = 0.05$.

5. Determine if the assumptions for homoscedasticity and normally distributed residuals are violated for the full linear model in (1). (Hint: the method of checking these assumptions is the same for multiple linear regression as in simple linear regression.)