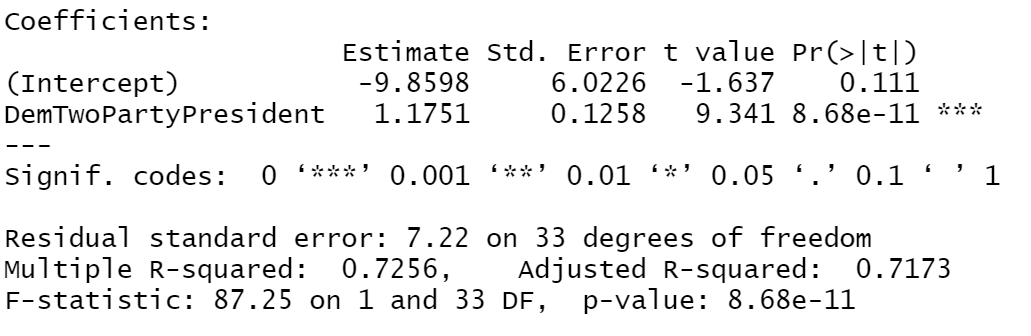
**Problem Set 7**

This problem set covers material from the two classes after the exam, including material from Chapter 8. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full. The Problem Set is due on Tuesday April 13 at 11:59 PM.

**Part I: 2020 Senate Elections**

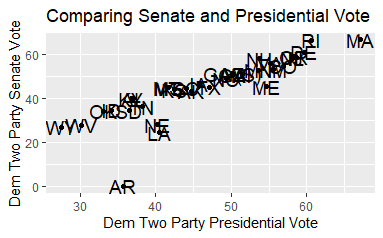
1. The first part of this problem set uses data from the 2020 Senate elections Below is a model where the dependent variable (response) the two-party percentage of vote Democratic Senate received in 2020 and the independent (explanatory) variable is the percentage of the two-party vote received by Joe Biden.



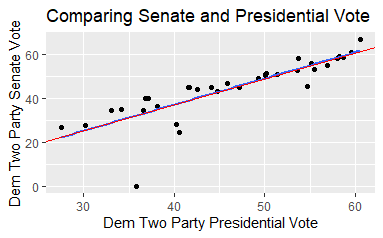
* 1. Please interpret the coefficients for the intercept and explanatory variable here and their associated p-values. (16 points)
  2. Please construct and interpret a 95% confidence interval around the coefficient for Democratic two-party presidential vote. What does this suggest about the statistical significance of this predictor. Please show your work. (8 pts.)
  3. Please interpret the R2 value for this regression. What does this R2 value suggest about how well 2020 presidential vote explains the variation in how well Democratic Senate candidates performed in a state? (8 points)
  4. Below is a scatterplot of the data and a residuals plot. Please assess each of the four requirements for the least squares line. (16 pts.)

|  |  |
| --- | --- |
|  |  |

* 1. Here is a scatterplot that includes state labels. Do there appear to be any outliers here? Should we just throw out these observation(s)? Do some research on Google to see why this/these points might be outliers. (10 pts.)



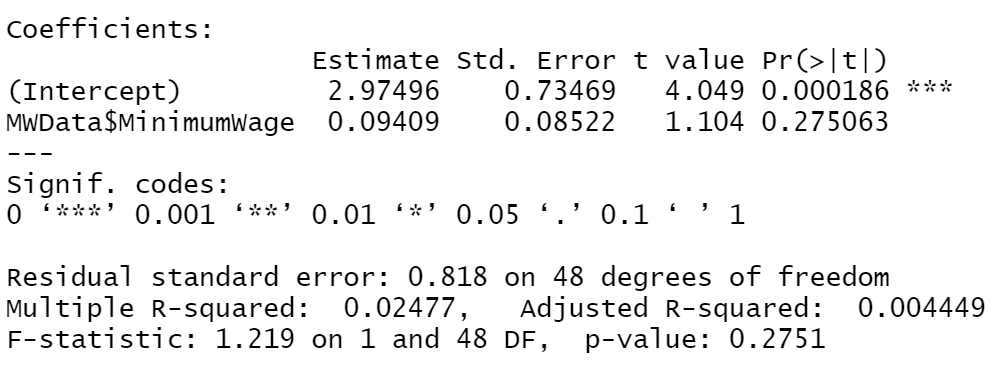
* 1. One data point (MA) had a larger percentage of Democratic two-party presidential vote than any other state. What kind of point could this potentially be based upon its location compared to other points? Below, I’ve included a scatterplot with the least squares line with all states included (red) and the least squares line for a model that omitted the MA observation (blue- yes, it’s there). Based upon this, should we be especially concerned about this point? (8 points)



* 1. Later this month, Congress will take up a bill to make Washington, D.C. a state. If Washington, D.C. becomes a state, it would elect 2 senators. In the 2020 election, Joe Biden received 94.46% of the two-party vote in Washington, D.C. Using this model, predict what percent of the vote a Democratic candidate for Congress would receive in Washington, D.C. if it were a state. What does this suggest about extrapolation? Please show your work. (8 pts.)

**Part II: Minimum Wage Levels and Unemployment**

1. A common argument used against raising the minimum wage is that it increases unemployment. Here, we are going to use data from 2018 to look at the relationship between minimum wage levels and unemployment in US states. Here is the output of an OLS regression model in R where unemployment was the dependent (response) variable and the minimum wage level in that state was the independent (explanatory) variable. States with a minimum wage equal or lower than the federal level are set at $7.25, the federal minimum wage in 2018. (26 pts.)



* 1. Please interpret the coefficient and p-value for the minimum wage variable and the R2 value for the model. Is there evidence here to suggest that higher minimum wage levels are associated with higher unemployment? What does this R2 value suggest about how well minimum wage levels explain variation in unemployment in 2018? (10 pts.)
  2. What is the correlation (Pearson’s r) between these two variables? Please show your work and explain your answer. Is this a strong correlation? (6 pts.)
  3. The standard deviation of the unemployment variable is 0.82. Using the information given or calculated earlier in this problem, what is the standard deviation in minimum wage levels? (6 pts.)
  4. Interpret the minimum wage coefficient using the one standard deviation increase framework. (4 pts.)