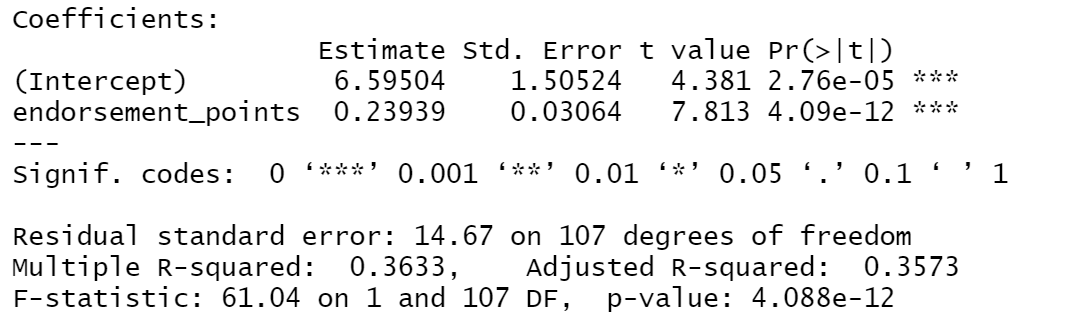
**Problem Set 7**

This problem set covers material from classes 21 and 22 including material from OIS sections 8.1 to 8.4. 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. This problem set is due on Gradescope by Wednesday November 10, 2021 at 11:59 PM.

**Part I: Bivariate Regression and the Invisible Primary (76 pts.)**

Before the actual voting begins in 2024, something called the “invisible primary” will take place, where presidential candidates will compete for endorsements. Here, we will look at data FiveThirtyEight collected on the invisible primary from 1980-2012 to learn a bit more about the role of endorsements and how they do (or do not) translate into voters.

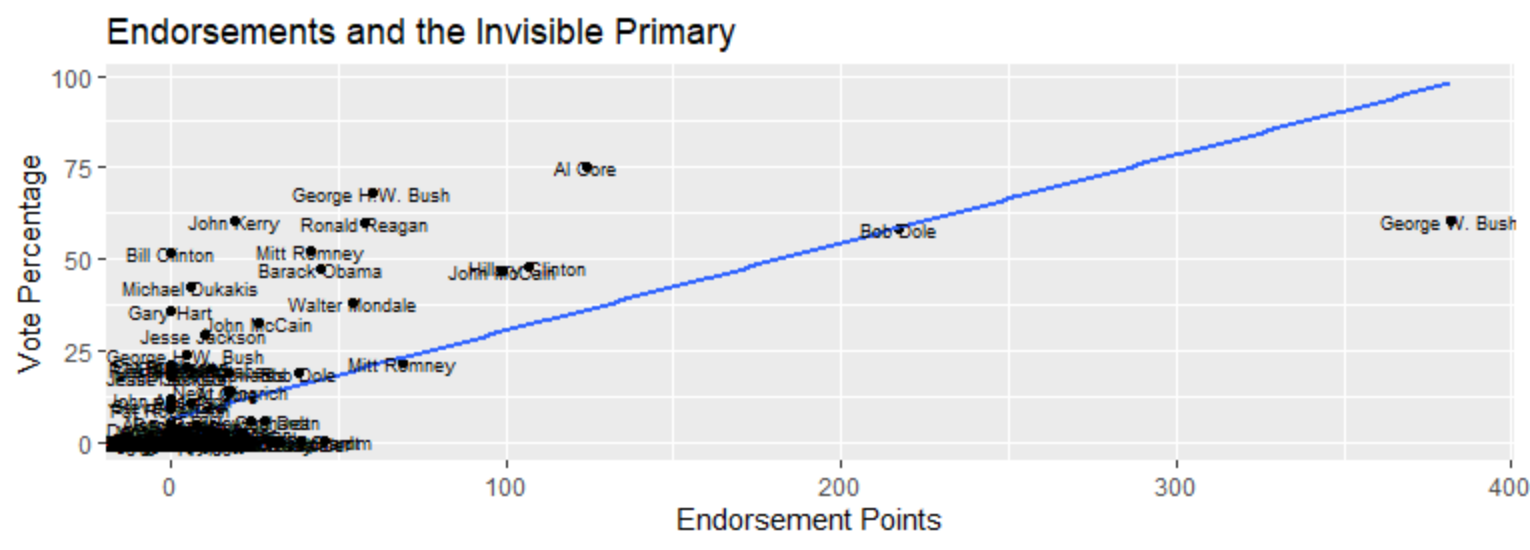
1. Below is a model where the dependent (response) variable is the percentage of vote received by the presidential candidate and the independent (explanatory) variable is the number of endorsements received by the presidential candidate. (76 pts.)



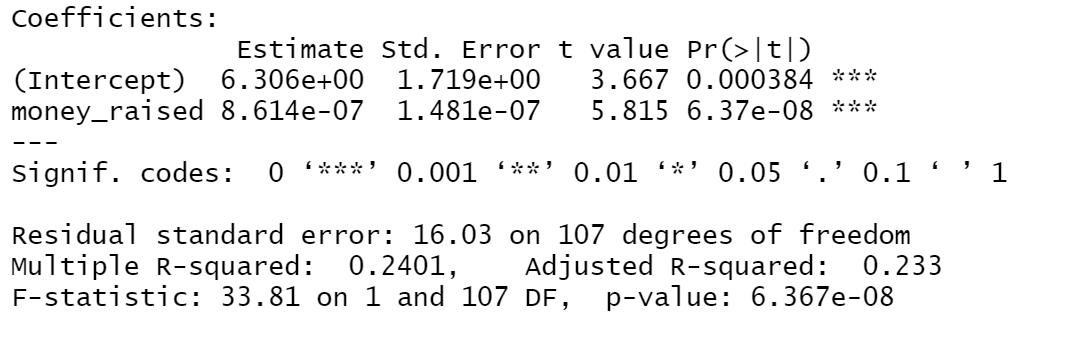
1. Please interpret the coefficients for the intercept and explanatory variable here and their associated p-values. (16 points)
2. Please construct and interpret a 95% confidence interval around the coefficient for endorsement points. What does this suggest about the statistical significance of this 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 endorsements explain the variation in primary vote percentage? (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 candidate name labels. Do there appear to be any outliers here? Should we just throw out these observation(s)? (8 pts.)

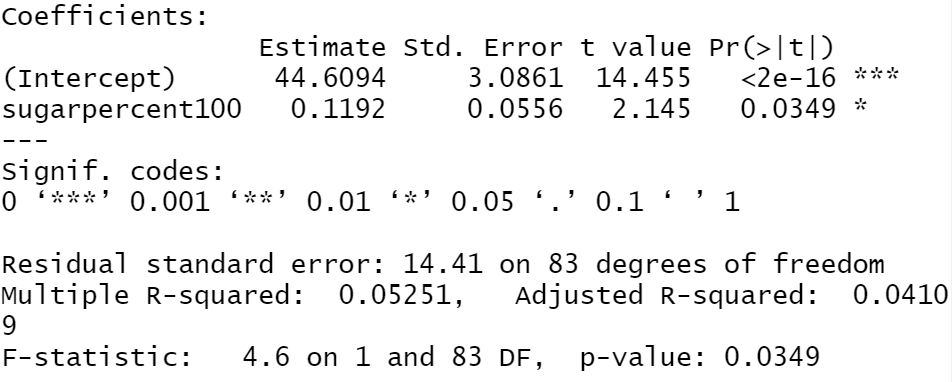


1. One data point, that for George W. Bush, has a greater value in endorsement points than any of the other here. What kind of point is this and what could it potentially be based upon its location compared to other points? How could we potentially assess if it is this type of point? (8 pts.)
2. In 2016, Hillary Clinton received 523 endorsement points, even more than George W. Bush in 2000. On average, what percentage of the vote would Clinton be predicted to receive? What does this number suggest about extrapolation? Please show your work. (8 pts.)
3. Does money move votes? Here is a regression where the dependent variable (response) the percentage of vote received by the presidential candidate and the independent (explanatory) variable is amount of money raised by the presidential candidate in dollars. Please comment on why this is such a “small” coefficient, despite being statistically significant, and what you might do to make it substantively easier to interpret. (4 pts.)



**II: Bivariate Regression and Candy Rankings (24 pts.)**

1. Are sugary candies more popular? Using the *FiveThirtyEight* Candy Rankings data set, here is a regression where the dependent variable (response variable) is the winning percentage and the independent variable (explanatory variable) is the sugar percentile. Both of these have been set to be on a scale of 0 to 100.



* 1. Please interpret the coefficient and p-value for the sugar percentile variable and the R2 value for the model. Is there evidence here to suggest that higher sugar content is associated with a higher winning percentage? What does this R2 value suggest about how well sugar content explains variation in winning percentage? (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? (4 pts.)
  3. The standard deviation of the winning percentage variable is 14.7. Using the information given or calculated earlier in this problem, what is the standard deviation in sugar content? (6 pts.)
  4. Interpret the sugar content coefficient using the one standard deviation increase framework. (4 pts.)