PSYC 5143
Spring 2024
Problem Set #11 (the last!)

This is due at 2:00 p.m. on Wednesday, May 1 on Blackboard, preferably in an R file.

The file cigarettes.csv has real data – for 496 senators and representatives who served in Congress in the late 1990s – one randomly selected vote for each politician on some tobacco-related legislation. (Why one vote? To get rid of any non-independence problems that might exist within persons.) The variables of interest here are *votedpro* (1 = voted in a way that was favorable to the tobacco industry; 0 = voted against) and *money*, a measure of how much money the legislator received from tobacco industry political action committees.

1) Fit a regular linear model with *votedpro* as the outcome and *money* as the predictor. Examine the status of the normality and homoscedasticity assumptions, and say for each how things look.

2) Generate a scatterplot with money on the x-axis and voting on the y-axis with a best-fitting line added, and a second one with a LOESS fit added. (Bonus points worth absolutely nothing if you can add both fit shapes to the same graph.) How do these leave you feeling about the linearity of the relationship between money and voting behavior? How do these leave you feeling about the usefulness of a linear model in making predictions about voting behavior? (It won't always be the case that a linear model is terrible for a binary outcome. Here's one defense of using linear models, at least under one set of circumstances. As with many things, it's complicated.)

3) Fit a logistic regression model with voting as the outcome and money as the predictor. Consider carefully whether to mean-center the predictor[1].
   a. Interpret the (logit) intercept and the slope in the model.
   b. Now exponentiate the intercept and slope and interpret what these (odds) values mean.
   c. Finally, convert the parameter estimates to probabilities and interpret what they mean.

[1] There are negative values. I don't know why. These are real data from Luke and Krauss (2004).